Mitsubishi Electric Research Laboratories (MERL)

Annual Report

July 2001 through June 2002

TR2002-00

Welcome to Mitsubishi Electric Research Laboratories (MERL), the North American corporate R&D arm of Mitsubishi Electric Corporation (MELCO). In this report you will find descriptions of MERL as a whole, our three laboratories (MERL Cambridge Research, MERL Cambridge Systems, and MERL Murray Hill), and most importantly our current projects.

201 Broadway, Cambridge, Massachusetts 02139
617.621.7500

Table of Contents

Mitsubishi Electric Research Laboratories

Mitsubishi Electric Research Laboratories (MERL) is the North American arm of the central research and development organization of Mitsubishi Electric Company (MELCO). MERL conducts application-motivated basic research and advanced development in computer and communications technology.

MERL's mission—our assignment from MELCO—is twofold.

- To generate highly significant intellectual property (papers, patents, and prototypes) in areas of importance to MELCO.
- To locate organizations within MELCO that can benefit from this technology and deliver it to them in forms they can use effectively (specifications, software components, and systems).

MERL's vision—our goal for ourselves—is also twofold.

- To be one of the world's premiere research laboratories, significantly advancing the frontiers of technology and making lasting impacts on the world.
- Within our areas of expertise, to be the prime source of new technology for MELCO.

MERL focuses on two key technology sectors:

Human/computer interaction (HCI) - featuring computer vision, computer graphics, speech interfaces and novel interaction devices

Digital communications - featuring video processing and networking.

The labs within MERL focus on specific segments of these technologies, while working collaboratively to achieve groundbreaking results. Our output ranges from papers and patents, through proof-of-concept hardware and software prototypes, to industry-first products.

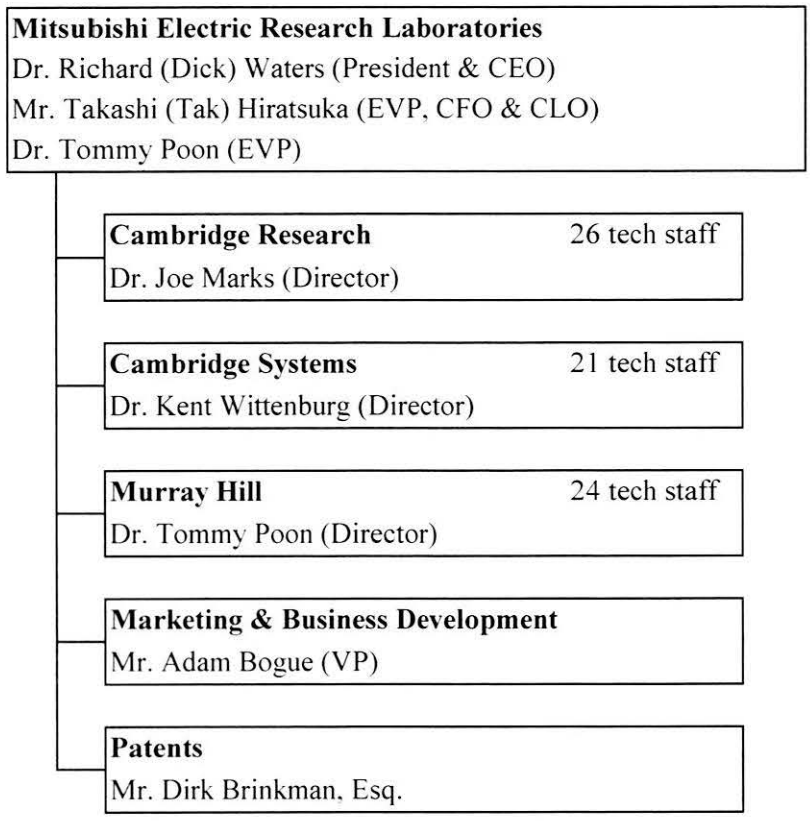
MERL is small enough to be agile and flexible in the dynamic marketplace of ideas. However, we gain leverage from the size, recognition and diversity of our strong global parent. We turn our technical achievements into business successes by partnering with MELCO's business units and with other labs in MELCO's global R&D network.

We are strongly involved in the R&D community and standards activities, maintaining long-standing cooperative relationships with a number of research universities including Brown, CMU, Columbia, ETH Zurich, Georgia Tech, MIT, Princeton and the University of Michigan. We encourage our staff to be involved in their professional communities via conferences, papers, and continuing professional development.

As shown in the chart on the next page, MERL consists of three laboratories, one in Murray Hill New Jersey and two in Cambridge Massachusetts. (See the lab description sections that follow.) MERL's headquarters (also in Cambridge) includes a small marketing and business development department to help assess the potential market impact of our work and an in-house patent department to speed the filing of patents.

This annual report is a snapshot of MERL's web site. For additional and updated information please visit "http://www.merl.com".

The following shows the basic organization of MERL and presents the management team.



Richard (Dick) Waters
Ph.D., Massachusetts Institute of Technology, 1978
President & Chief Executive Officer; Research Fellow Cambridge Research

Dick Waters received his Ph.D. in artificial intelligence (AI). For the next 13 years he worked at the MIT AI Lab as a Research Scientist and co-principal investigator of the Programmer's Apprentice project. Dick was a founding member of MERL Cambridge Research in 1991. As a MERL researcher his work centered on multi-user interactive environments for work, learning and play. In January 1998, Dick became Director of MERL Cambridge Research. In December 1999, he became CEO of MERL as a whole.

Takashi (Tak) Hiratsuka
B.S., Univ. of Tokyo, 1969
Executive Vice President, Chief Financial Officer & Chief Liaison Officer

After many years in MELCO's computer business unit, Tak Hiratsuka came to the US in 1988, as the liaison between MELCO and MELCO's subsidiary Horizon Research Inc. After 6 years there, Tak took on other duties for MELCO including a stint as Assistant Deputy Director of Corporate R&D in Tokyo. In December 1999, Tak returned to the US as MELCO's Liaison to MERL. Tak plays an essential role in building partnerships between MERL and MELCO.

Tommy Poon
Ph.D., Columbia University, 1980
Executive Vice President; Director Murray Hill

After receiving his Ph.D., Tommy Poon worked briefly at RCA Laboratory before moving to AT&T Bell Laboratories. During 13 years with Bell Labs, Tommy managed numerous R&D projects in telecommunications and signal processing. He joined MERL Murray Hill as Director in 1995. His primary research interests include digital & wireless communications, digital video and digital networks.

Joe Marks
Ph.D., Harvard University, 1991
Vice President; Director Cambridge Research

Prior to Joining MERL in 1994, Joe Marks worked at Digital Equipment Corporation's Cambridge Research Laboratory. As a researcher at MERL, Joe's primary focus was on computer graphics, user interfaces, and heuristic optimization. Joe also has a strong interest in teaching. He was an adjunct lecturer in the Division of Engineering and Applied Sciences at Harvard University from 1991 to 1996. Joe became associate director of MERL Cambridge Research in 1999 and director in 2000.

Kent Wittenburg
PhD., University of Texas at Austin, 1986
Vice President; Director Cambridge Systems

Before Joining MERL in 2001, Kent worked at the Microcoelectronics and Computer Technology Corporation (MCC), Bellcore and then Verizon/GTE laboratories. He has done research on many aspects of Human/Computer Interaction including multidimensional information visualization, visual languages for diagrams, and natural language parsing. Kent become director MERL Cambridge Systems in 2002.

Adam Bogue
B.S., MIT, 1986; MBA, MIT Sloan School, 1990
Vice President of Marketing and Business Development

For 3 years at GenRad Inc, Adam Bogue was responsible for managing a new line of automatic test equipment. Prior to joining MERL, Adam spent 7 years at Active Control eXperts Inc. beginning as Director of Sales and Marketing and ending as Vice President, Core and New Business Unit. Adam began work for MERL in June of 2000.

Dirk Brinkman, Esq.
M.Sc., Univ. of Toronto, 1970; J.D., Suffolk Univ. Law School, 1990
Patent Counsel

Dirk Brinkman's undergraduate and Masters work was in Medical Physics. Prior to joining MERL in 1998, he spent most of his career at Digital Equipment Corporation, first as an engineer and product manger in the Medical Systems Group, and then as a Patent Attorney for Digital's Research Laboratories in Cambridge MA and Palo Alto CA.

Mitsubishi Electric

Number 141 on Fortune magazine's 2002 list of the world's 500 largest corporations, Mitsubishi Electric Corporation (MELCO) has approximately $29 billion in annual sales and 117,000 employees in 34 countries. Like most Japanese companies, the lingering malaise of the Japanese economy coupled with the worldwide slump in semiconductors and cell phones lead to losses in 2001. However, the combination of strong steps at MELCO to improve its business position and a reviving world economy in 2002 should lead to renewed profits.

MELCO is organized into nine business units (listed below in order of revenues). Because information technology is important to each of these business units, MERL works with them all. (The rightmost column below shows the abbreviated Japanese business unit names commonly used by MELCO insiders.)

Mitsubishi Electric Diversified Electrical and Electronics Manufacturer		MELCO
Living Environment & Digital Media Equip. Air Conditioners, Refrigerators, TVs, DVDs, LCD Projectors	(Shizuoka, Kyoto)	Lihon
Semiconductors Memory, Microcontrollers, Optical Semiconductors, LCD Panels	(Kita Itami)	Hanpon
Building Systems Elevators, Escalators, Building Monitoring	(Inazawa)	Biruhon
Social Infrastructure Systems Power Equipment, Plant Control, Transportation	(Kobe, Itami)	Shakaihon
Electronic Systems Satellites, Radar, Military Systems	(Kamakura, Itami)	Denshihon
Communication Systems Wired Communications, Broadcast Communications, Cell Phones	(Kamakura, Itami)	Tsuhon
Automotive Equipment Alternators, Engine Controllers, Car Stereos, Car Navigation	(Himeji, Sanda)	Shahon
Factory Automation Programmable Logic Controllers, Industrial Machine Tools	(Nagoya)	FAhon
Information Systems and Services Turnkey Information Systems, Computer Hardware, Networking Services	(Tokyo, Kamakura)	IShon

It is worthy of note that there are over 30 major independent companies that use the word Mitsubishi in their names. These companies include the Mitsubishi trading company, Mitsubishi Chemical, Mitsubishi Heavy Industries, Mitsubishi Motors and Mitsubishi-Tokyo Financial Group (all five of which are also on the Fortune Global 500 list). They have shared roots in 19th century Japan; however, these companies have been separate for many years and MELCO has been separate from all of them since its founding in 1921.

Many of MELCO's business units have North American subsidiaries. These subsidiaries, particularly those that do design as well as manufacturing and sales, are natural partners of MERL. The subsidiaries that have sales of over one hundred million dollars per year are listed below in order of sales volume.

Mitsubishi Digital Electronics America, Inc. (MDEA)
Design, Manufacturing & Sales: Lihon (Los Angeles, Mexicali MX)
High Definition Projection Televisions, DVDs, VCRs

Mitsubishi Electric United States, Inc. (MEUS)
Sales: Several BUs (Los Angeles, Sunnyvale & other cities)
Semiconductors, Air Conditioning, Elevators

Mitsubishi Electric Automotive America, Inc. (MEAA)
Manufacturing & Sales: Shahon (Detroit, Mason OH)
Auto Parts

Mitsubishi Electric Power Products, Inc. (MEPPI)
Design, Manufacturing & Sales: Shakaihon (Pittsburgh)
Power Transmission Products

Mitsubishi Electric Sales Canada, Inc. (MESCA)
Sales: Several BUs (Toronto & other cities)
Semiconductors, Air Conditioning, Projectors

Mitsubishi Electric Display Devices America, Inc. (MDDA)
Manufacturing & Sales: Lihon (Calexico CA, Mexicali MX)
Cathode Ray Tubes for computer monitors

Mitsubishi Electric Automation, Inc. (MEAU)
Sales & Installation: FAhon (Chicago)
Factory Automation Equipment

Mitsubishi Electric Corporate R&D

MERL is one of five laboratories in MELCO's Global Corporate R&D network. The chart below summarizes the primary activities of these labs. MERL pursues collaborations with all these labs. (The rightmost column shows the Japanese nicknames commonly used by insiders.)

Corporate R&D Headquarters: Dr. H.Ogata (director), Dr. H.Koezuka (GM), 20 people (Tokyo) Managing MELCO's R&D	Hatsuhon
Advanced Technology R&D Center (ATC) Research & Advanced Development: Dr. T.Yamada (GM), 1030 people (Itami) Materials, Semiconductor Devices, Electrical & Mechanical Engineering	Sentansoken
Information Technology R&D Center (ITC) Advanced Development: Mr. S.Ono (GM), 1000 People (Ofuna) Information Systems, Communications, Opto-Electronics	Johosoken
Industrial Design Center (IDC) Advanced Development: Mr. K.Chiba (GM), 100 people (Ofuna) Industrial Design, Usability Studies	IDken
Mitsubishi Electric Research Laboratories (MERL) Research & Advanced Development: Dr. R.Waters (CEO), 90 people (MA & NJ) Computer Vision, HCI, Internet Software, Digital Audio & Video Communications	MERL
Mitsubishi Electric Information Technology Centre Europe (ITE) Advanced Development: Mr. R.Nishii (CEO), 60 people (France & England) Wireless Communications, Digital Audio & Video Broadcasting	ITE

MERL Cambridge Research

MERL Cambridge Research pursues basic research in applied computing. Our efforts are directed towards applications of practical significance, but our time horizon is long (five or more years) and our appetite for technical challenge and risk is high. Located in Cambridge, Massachusetts, hometown of Harvard University and the Massachusetts Institute of Technology, the laboratory currently has a technical staff of 26. The permanent staff is enriched by an active program of student internships hosting approximately 40 students per year for an average of three to four months each.

The best ideas are born and mature most quickly when critically examined and refined by many minds. Toward this end, MERL researchers are encouraged to work together with each other and with researchers at other institutions. In particular, we forge collaborations that carry our technologies into products by connecting to our sister organizations in MERL and MELCO. Further, we participate actively in external research communities, exposing our work to critical peer review, and publishing our results as quickly as possible.

To allow rapid response to opportunities, MERL Cambridge Research is organized as a flexible community of researchers without internal divisions. Our goal is to support a continually changing mix of individual explorations and group projects, where promising individual efforts can easily grow into projects and projects can easily disband at the end of their natural lives. Each researcher is typically involved in both individual explorations and related group projects.

Over the last few years, perceptual user interaction (PUI) has emerged as a central theme of the lab. This theme is providing a unifying focus for our research in computer graphics, computer vision, audio and image processing, natural language processing, speech processing, machine learning, networked and wireless communication, and human-computer interaction. Concretely, this has generated projects in such important business areas as video surveillance, telephony, and consumer electronics. The PUI theme is resonating well with MELCO labs and business units, producing better connections for our research than ever before.

Yet while we are strengthening our connections with MELCO's business interests, we are also improving our record of scientific and technological achievement, as measured by our recent publications and standards activities. Within its areas of expertise, MERL Cambridge Research is widely regarded as one of the premier industrial research laboratories in the world.

Successful basic research in an industrial setting requires many things: a world-class team of researchers; a long-term commitment to open-ended, risky exploration; research themes that resonate well with the parent company; and effective tech-transfer mechanisms whereby research ideas can be converted into new products. These elements are all in place at MERL Cambridge Research.

MERL Cambridge Systems

MERL Cambridge Systems is co-located with MERL Cambridge Research and its programs are deliberately intertwined. Where MERL Cambridge Research is composed largely of individual researchers with a time horizon of five or more years, MERL Cambridge Systems is devoted to the practical realization of technology innovations in a one to four year time frame. This goal requires group structure and coordination, and thus its organization is composed of teams with typically 4-6 members. These teams produce early concept prototypes alongside MERL Cambridge Researchers, but more often than not the goal is to bring the research to the next stage of development, solving the myriad of technical issues that stand in the way of delivering a gleam in a researcher's eye into a full-bodied form with business impact. The lab is divided into four technology areas: Vision Applications, Speech Applications, eServices, and Human-Computer Interaction and Devices.

Vision Applications is devoted to building software libraries and systems for a variety of applications that make use of the world-class computer vision research emerging from MERL Cambridge Research. These applications include security, safety of elevator operations, analysis and enhancement of imagery, and demographic analysis from video data. This group leads a multi-year project funded from MELCO Central R&D that is focused on building a software platform that is aligned with business units in the area of computer-human observation.

Speech Applications focuses on finding innovative applications of speech technology. While the group does not develop core speech engines, it does develop the modules and applications that are built around speech engines. The group is fortunate to have staff and leadership with more than 20 years of collective experience developing actual products in the industry including core speech engines. Among this year's projects are an application of speech in document query systems as well as a collaboration with a MELCO industrial design group who have developed innovative prototypes of speech applications in the automotive industry. The group is also contributing to a multi-year project led by a Japanese sister lab whose goal is to develop a web-based infrastructure for speech applications.

The E-services group has had a successful history of developing mobile agent software that has figured prominently in the plans of the MELCO Social Infrastructure Systems business unit. Work continues in this Concordia project on the security layer this year. Another long-term success story has been the NetRep project, which continues to develop new features for a commercially successful system for network data replication. E-services also has new initiatives in the areas of data analysis and visualization. The group worked with CRL this year on a realtime recommender system and has demonstrated how visualization software can enhance its value.

Human-Computer Interaction and Devices focuses on innovative devices and their interaction off the desktop. The group's flagship project is the DiamondTouch table, which includes the multi-user touch technology hardware itself as well as software drivers and a new system development kit. Among the group's new devices this year are iGlassware and ultra-cheap two-way LED communication. HCI and Devices also had a project applying human-guided search and visualization technologies to the problem of antenna design as well as an early foray into the area of multimedia interfaces with engaging 3D interactive presentations of video frames.

MERL Murray Hill

MERL Murray Hill is structured to address four major areas of research and advanced technology development: Advanced Communication Technologies, Audio and Video Systems, Digital Networks, and Visual Communications.
The Advanced Communication Technologies (ACT) group focuses on high speed digital communications systems with high quality of services (QoS) including: next generation wireless communications systems, wireless local area networks, high speed Ethernet and ultra broadband technologies. ACT creates new concepts and technologies, new algorithms and new designs that enhances MELCO's IPR portfolio. ACT participates in industry standards development, such as 3GPP, ITU-T, ITU-R, and IEEE802.3ae, 11, 15, and 16. The team is also collaborating with MELCO's business units to develop competitive communications products.

The Audio and Video Systems (AVS) group develops innovative techniques of multimedia content analysis for diverse applications. AVS research projects focus on multimedia storage and retrieval as well as multimedia content analysis. The key techniques include rapid audio/video browsing, indexing, summarization, event detection, content mining and video segmentation. The team has been actively and contributing to the MPEG standards, especially to the MPEG-7.

The Digital Networks (DN) group focuses on the design and integration of key technologies to link broadband and narrowband external networks to local (and/or personal) networks based on wireless and wireline physical layers. By taking advantage of key open protocols such as TCP, MPEG, IEEE1394, and others; a wide variety of end-to-end usage models supporting QoS, plug-and-play, guaranteed delivery, mobility, and security are being investigated. DN is actively participating in various industry standards related to Home Networking, Digital Broadcasting, and Consumer Electronics such as HAVi, 1394TA & UPnP, ATSC, EIA CEA and ATTC.

The Visual Communications (VC) group is developing breakthrough ideas related to the communication and exchange of visual information. VC projects focus on the current and future interests of MELCO labs and BU's which include the development of video coding algorithms, an improved HDTV receiver, video transcoding algorithms, Universal Multimedia Access (UMA) and video streaming systems. The team is actively participating and contributing to the development of MPEG and other standards, including MPEG-2, MPEG-4, MPEG-7 and MPEG-21, JVT and IETF.

MERL Murray Hill Lab has filed more than 90 patents and has published more than 100 papers arising from the above activities. Its expertise includes: 1) System knowledge of digital video-coding algorithms, multimedia content processing/analysis, digital communications systems, wireless & wireline communications, digital broadcasting and home-networking; 2) Active participation and contributions to U.S. & international standards; and 3) Fast prototyping of real-time hardware & firmware and digital signal processing.

MERL Murray Hill Lab has developed close working relationships with various MELCO Business Units and R&D Labs including: Mitsubishi Digital Electronics America - MDEA (digital projection TV with 1394 interfaces), the Information Technology R&D Center - Johosoken (digital video technology & MPEG standards), and the Advanced Technology R&D Center - Sentansoken (surveillance systems, DVD recorders & video content analysis), the System LSI Development Center - L-Ji-Se (wireless LAN, home networking & 10 Gbps Ethernet). The lab is located in Murray Hill, New Jersey, with easy access to New York City.

Technical Staff

Paul Beardsley
Ph.D., Oxford University, 1992
Research Scientist, MERL Cambridge Research

Paul Beardsley is a computer vision researcher. He obtained a DPhil from Oxford for work in applications of projective geometry to computer vision. He is currently working on a geometry-based vision library in C++. Current projects are on stereo-vision for surveillance, and on a hand-held 3D scanner.

Ghulam Bhatti
Ph.D., Boston University, 1998
Member Technical Staff, MERL Murray Hill

Ghulam Bhatti received his Ph.D. in Computer Science, specializing in Distributed and Parallel Discrete Event Simulation. He joined MHL in November 2000. Previously, Ghulam has worked as a Sr. Software Engineer at Evare LLC, Inc, developing software for a network switch, named as Winternet, and implementing an RSA cryptographic scheme. He also worked at Excel Tech. Ltd. (XLTEK) as an Embedded Software Engineer, developing embedded software for a portable EEG device. Currently, Ghulam is working on Home Networking and Digital TV. His interests include algorithms, embedded software development, and networking.

Emmanuelle Bourrat
M.S., Université de Rennes1, 1999
Member Technical Staff, MERL Cambridge Systems

Emmanuelle Bourrat graduated in 1999 from IRISA (Institut de Recherche en Informatique et Systèmes Aléatoires), Université de Rennes1, France. She earned a Master's degree in computer science and digital images. She was an intern at MERL from February 2000 to August 2000, where she worked on a tool to visualize and segment scans of the head. She joined CSL as a full-time employee in August 2000. Her main interest is in computer vision and medical imaging. She is currently working on the CHO (Computer Human Observation) project.

Matthew Brand
Ph.D., Northwestern University, 1994
Research Scientist, MERL Cambridge Research

Matthew Brand's work focuses on unsupervised learning from sensory data. One goal is to make machines that learn to realistically mimic and augment human performances. Recent results include an entropy optimization framework for learning, recovery of nonrigid 3D shape from ordinary video, a linear-time online SVD, and realistic synthesis of human and nature scenes from probability models. Brand has been named one of 100 top innovators of his generation (Technology Review 1999) and one of industry's "top R&D stars" (Industry Week 2000). Recent academic honors include best computer vision paper award (CVPR 2001).

Michael Casey
Ph.D., Massachusetts Institute of Technology, 1998
Research Scientist, MERL Cambridge Research

Michael Casey is conducting research on machine learning systems for speech, music and general audio. Applications include generalized sound recognition, speaker identification using low-complexity speech models, music retreival by similairty and speech enhancement in noisy environments. Michael is an editor and co-chair for the MPEG7 international standard for which he has developed new technologies in generalized sound recognition and similarity and he has co-ordinated the implementation of the audio reference software. In addition to his research, Michael holds two international awards for electronic music composition. In 1997 Michael was nominated as a finalist in the Discover Awards for Technological Innovation for his research on structured audio. He holds a US patent for generalized sound recognition.

Johnas Cukier
M.Sc., Polytechnic Institute of New York, 1985
Principal Member Technical Staff, MERL Murray Hill

Johnas Cukier received his B.Sc. degree in Physics and Computer Science in 1983 from New York University and his M.Sc. degree in Electrical Engineering in 1985 from Polytechnic Institute of New York. He joined MERL Murray Hill in 1996, working on digital systems for CATV, RF microwave transmitters/receivers, and front-end advanced TV receivers. His current interests are in advanced Digital Networking and Digital Signal Processing.

Andrew J. Curtin
J.D., Suffolk University Law School, 1997
Associate Patent Counsel, MERL Headquarters

Andrew Curtin received his B.S. in Marine Engineering from the Massachusetts Maritime Academy and his J.D. from Suffolk University Law School. Prior to joining MERL in 2001, he spent six years as an engineering officer aboard U.S. flag merchant ships engaged in world-wide trade, followed by two years as an attorney in private practice.

Paul Dietz
Ph.D., Carnegie Mellon University, 1995
Principal Technical Staff, MERL Cambridge Systems

Paul Dietz received his B.S. in Electrical Engineering from MIT, the M.S. and Ph.D. from Carnegie Mellon University. Prior to joining MERL, Paul headed up the electrical engineering efforts at Walt Disney Imagineering's Cambridge R&D lab where he worked on a wide variety of projects including theme park attractions, systems for the ABC television network and consumer products. He joined MERL Cambridge Systems in 2000 and has been leading efforts developing new user interface technologies.

Ajay Divakaran
Ph.D., Rensselaer Polytechnic Institute, 1993
Principal Member Technical Staff, MERL Murray Hill

Ajay Divakaran received the B.E. egree in Electronics and Communication Engineering from the University of Jodhpur, Jodhpur, India, in 1985, and the M.S. and Ph.D. degrees from Rensselaer Polytechnic Institute, in 1988 and 1993 respectively. He worked as a research associate at the Indian Institute of Science before joining Iterated Systems Inc., Atlanta, GA in 1995. At Iterated Systems he worked on video-coding algorithms for video telephony and entertainment-quality video. In 1998 he joined MERL Murray Hill, where he has worked on video indexing and summarization with a view to MPEG-7 applications.

Alan Esenther
M.Sc., Boston University, 1993
Principal Technical Staff, MERL Cambridge Systems

Alan Esenther (M.S.C.E) enjoys Internet technologies, human-computer interaction, distributed software development, dynamic web programming, GUI work, and usability enhancements. He has taken an interest in finding easier ways to accomplish software tasks -- ideally leveraging existing end-user resources. Recent work involves instant co-browsing (lightweight real-time distributed collaboration using unmodified web browsers). Previous work includes interactive page generation for mobile agents, an email gateway for offline surfing, transaction monitors, kernel-level volume management, and microprocessor development.

George Fang
B.Sc., California Institute of Technology, 1990
Member Technical Staff, MERL Murray Hill

George Fang received his B.Sc. degree from California Institute of Technology and became a member of Mitsubishi Electric Corporation's Kyoto Works in 1990. During the ten years working in Japan, he was a hardware engineer designing analog and digital consumer televisions for the American market and coordinated joint design efforts between Japan and the United States. He joined the Murray Hill Laboratory in February of 2001 with research objectives in wireless and network technologies.

James Fang
B.Sc., Columbia University, 1992
Member Technical Staff, MERL Murray Hill

James Fang received his B.Sc. from Columbia University in 1992 and did some graduate work there before joining Mitsubishi Electric in 1995. He worked on consumer televisions for three years before transferring to MERL Murray Hill in 1998. He is currently working on digital wireless communications.

Sarah Frisken
Ph.D., Carnegie Mellon University, 1991
Senior Research Scientist, MERL Cambridge Research

Sarah Frisken (formerly Gibson) has research interests in computer graphics, volume visualization and physically based modeling. She has led a team of researchers and students to build a knee arthroscopy simulator that incorporates high-quality rendering, haptic feedback and physical modeling to simulate interactions between surgical tools and a computer model derived from 3D MRI data. Her current work is with Adaptively Sampled Distance Fields, a general representation of shape for computer graphics, which provides intuitive manipulation, and editing of smooth surfaces with fine detail. Applications include digital sculpting, volumetric effects for rendering, color gamut representation, path planning for CNC milling, and rapid prototyping.

Andrew Garland
Ph.D., Brandeis University, 2000
Visiting Scientist, MERL Cambridge Research

Andrew Garland is interested in artificial intelligence issues concerning agents that reason and learn. Recently, he developed a framework that justifies an agent's goal-directed behavior even in the absence of a provably correct plan. He has also authored, and implemented, techniques for a computer to work with a human in order to develop a hierarchical task model from examples. Such a model can be used, among other things, to represent shared beliefs about how an agent and a human can collaborate to accomplish tasks. At Brandeis, his thesis was on memory-based techniques for autonomous agents to learn how to better coordinate joint activities.

Daqing Gu
Ph.D., SUNY Stony Brook, 1999
Principal Technical Staff, MERL Murray Hill

Daqing Gu received the BE degree from Tsinghua University, Beijing, China in 1987; the MS and Ph.D. degrees in electrical engineering from the State University of New York at Stony Brook, Stony Brook, NY in 1996 and 1999, respectively. He joined MERL Murray Hill in 1999, and is currently a principal member of technical staff. His research interests include wireless communications, wireless and optical networking, QoS for WLAN and wireless communications.

Jianlin Guo
Ph.D., Windsor University, 1995
Principal Technical Staff, MERL Murray Hill

Jainlin Guo received his Ph.D. from Windsor University in 1995. He worked at Waterloo Maple for a year and a half as a software developer and joined MERL Murray Hill in 1998. He has published seven research papers and his primary research interests include home networks, digital broadcasting, and wireless computing.

Bret Harsham
Massachusetts Institute of Technology
Principal Technical Staff, MERL Cambridge Systems

Bret Harsham joined MERL Cambridge Systems in January 2001 to pursue interests in speech user interfaces and speech-centric devices. Prior to joining MERL, Bret spent 3 1/2 years at Dragon Systems designing and implementing handheld and automotive speech products. Earlier, he was a principal architect of a Firewall and Virtual Private Network product. Bret's other technical interests include distributed architectures, knowledge representation and language theory.

Jyhchau (Henry) Horng
Ph.D., Polytechnic University, 1998
Member Technical Staff, MERL Murray Hill

Henry Horng received the Ph.D. from Polytechnic University in 1998. He has worked as a research assistant at Polytechnic and as software developer and lecturer for Chung Cheng Institute of Technology, Taiwan. Henry joined MERL Murray Hill in 1999. His primary research interests include digital signal processing and communications.

Frederick J. Igo, Jr.
B.A., LeMoyne College
Senior Principal Technical Staff, MERL Cambridge Systems

Fred Igo's professional interests are in software development and its process. Starting at IPL Systems and continuing at Horizon Research Inc. (HRI), Mitsubishi Electric Information Technology Center America (MEITCA) and now MERL Cambridge Systems, Fred has been working with Mitsubishi Electric for about 20 years. During that time he has worked on various software technology, including Distributed Computing, Distributed OLTP, Message Queuing, Mobile Agents, OLAP/MDDB and Data Mining.

Michael Jones
Ph.D., Massachusetts Institute of Technology
Principal Technical Staff, MERL Cambridge Systems

Mike Jones joined MERL in the fall of 2001 after 4 years at the Digital/Compaq Cambridge Research Laboratory. Mike's main area of interest is computer vision, and he is particularly interested in using machine learning approaches for solving computer vision problems. He has focused on algorithms for detecting and analyzing people in images and video such as face detection, skin detection and facial analysis using morphable models. Recent Projects include Fast Face Detection using a Cascade of Detectors.

Hao-Song Kong
Ph.D., Sydney University, 1998
Member Technical Staff, MERL Murray Hill

Hao-Song Kong worked at Motorola Australian Research Center over 4 years before joining MERL Murray Hill Research Lab in 2001. He has research interests in neural networks, digital signal processing, image processing, video transcoding, video transmission and networking. His current projects are DVD recording, real-time video streaming platform development and QoS provisioning research for the next generation wired and wireless IP networks.

Darren Leigh
Ph.D., Harvard University, 1998
Research Scientist, MERL Cambridge Research

Darren Leigh's research interests range from electronic hardware and communications to operating systems and signal processing. Before coming to MERL Cambridge Research he worked on the Harvard University/Planetary Society Billion-channel ExtraTerrestrial Assay (Project BETA), a search for microwave signals from extraterrestrial civilizations. Other previous research includes 3D microscopic scanning, desktop manufacturing and network architectures for multimedia. His current research includes the Personal Eyewitness and interfacing and applications of the M32R/D and Artificial Retina chips.

Neal Lesh
Ph.D., University of Washington, 1998
Research Scientist, MERL Cambridge Research

Neal Lesh's research efforts aim to improve (or at least ease) the interaction between people and computers. His research projects include interactive optimization (the HuGS project), collaborative interface agents (the COLLAGEN project), and collaborative navigation of digital data (the Personal Digital Historian project). Before coming to MERL Cambridge Research, he was a graduate student at the University of Washington with Oren Etzioni, and and a postdoc with James Allen at the University of Rochester.

Sergei Makar-Limanov
Ph.D., Stanford University, 1994
Principal Technical Staff, MERL Cambridge Systems

Sergei Makar-Limanov received his Bachelors degree from University of Chicago, and his PhD in Mathematics from Stanford University. After spending a few years in academia, Sergei decided to join the "real world" where he worked on such software projects as computer aided manufacturing for PTC corporation and supply management for Kewill PLC. His most recent work involved designing automated software scalability testing tools at Empirix Corporation. Sergei has joined MERL in May 2001 to work on the Concordia project. He is now working with the vision group on creating a database framework for video surveillance applications.

Fernando Matsubara
M.Sc., University of Tokyo, 1990
Senior Principal Member Technical Staff, MERL Murray Hill

Fernando joined MELCO in Japan after graduating from the University of Tokyo (M.Sc.) He was assigned in the US as Manager of New Technology (1995) and Director of Technology (1997) for Mitsubishi Digital Broadcasting America before moving to MHL in 2000. He lead intra-company evangelism to adopt a digital connectivity solution for DTV, CE, and Gateway devices. He contributed to two key digital networking standards for Digital TV: OpenCable and CEA R4. Currently he is the Chairman of the HAVi Technical Steering Committee. Primary research areas are digital communications, network technology, and digital video.

David McDonald
Ph.D., Massachusetts Institute of Technology, 1980
Principal Technical Staff, MERL Cambridge Systems

David McDonald is a computational linguist. He pioneered the field of generation of natural language texts from a knowledge representation. After getting his degree in Artificial Intelligence from MIT he was on the Computer Science faculty at UMass Amherst where he directed the work of seven Ph.D. students in problems from tutoring to domain modeling to typography. After working at several natural language companies he has recently joined MERL as a member of the COLLAGEN project where he works at the interface between the symbolic reasoning of Collagen and the probabilistic reasoning of speech understanding.

Baback Moghaddam
Ph.D., Massachusetts Institute of Technology, 1997
Research Scientist, MERL Cambridge Research

Baback Moghaddam's research is in computational vision with focus on probabilistic visual learning, statistical modeling and pattern recognition with applications in biometrics and computer-human interfaces. Prior to MERL, Dr. Moghaddam was at the Vision and Modeling Group at the MIT Media Laboratory where he developed an automatic vision system that won DARPA's 1996 "FERET" face recognition competition. His previous research included fractal image compression, segmentation of synthetic aperture radar (SAR) imagery as well as designing a zero-gravity experiment for laser annealing of amorphous silicon which was flown aboard the US space shuttle in 1990.

Yves-Paul Nakache
M.Sc., E.S.I.E.E., 2000
Member Technical Staff, MERL Murray Hill

Yves-Paul Nakache received a French Engineering diploma equivalent to M.Sc. degree in Electrical Engineering in 2000 from E.S.I.E.E. (Ecole Supérieure d'Ingénieurs en Electrotechnique et Electronique), Paris, France. He joined MERL Murray Hill in 2000, where he is currently a Member of the Technical Staff. He works on Interference Cancellation and 3G CDMA systems. His current interests are in speech processing and wireless communications.

Philip Orlik
Ph.D., SUNY Stony Brook, 1999
Member Technical Staff, MERL Murray Hill

Philip Orlik received the B.E. degree in 1994 and the M.S. degree in 1997 both from the State University of New York at Stony Brook. In 1999 he earned his Ph. D. in electrical engineering also from SUNY Stony Brook. He joined MERL Murray Hill in August 2000, and is currently a member of technical staff. His research interests include wireless and optical communications, networking, queuing theory, and analytical modeling.

Kadir Peker
Ph.D., New Jersey Institute of Technology, 2001
Member Technical Staff, MERL Murray Hill

Kadir A. Peker received the B.S. degree from Bilkent University, Turkey in 1993, the M.S. degree from Rutgers University in 1996, and the Ph.D. degree from New Jersey Institute of Technology in 2001, all in Electrical Engineering. His Ph.D. dissertation is on content based video indexing and summarization using motion activity. He worked at MERL Murray Hill as an intern for more than a year, and joined as a member of technical staff in 2000. He has contributed to MPEG-7, published conference papers, submitted journal papers and filed patents on his Ph.D. topic. His current research interests include video indexing and summarization, home networking and multimedia networks.

Ron Perry
B.Sc., Bucknell University, 1981
Research Scientist, MERL Cambridge Research

Ron Perry joined MERL Cambridge Research as a Research Scientist in 1998. Prior to that, he was a consulting engineer at DEC developing a three-dimensional rendering ASIC called Neon. Ron has consulted for many companies including Kodak, Atex, Adobe, Quark, and Apple over the last 20 years, developing software and hardware products in the areas of computer graphics, imaging, color, and desktop publishing. Ron's research interests include fundamental algorithms in computer graphics with occasional excursions in numerical analysis and protein folding.

Hanspeter Pfister
Ph.D., State University of New York at Stony Brook, 1996
Associate Director and Senior Research Scientist, MERL Cambridge Research

Hanspeter Pfister is Associate Director and Senior Research Scientist at MERL Cambridge Research. He is the chief architect of VolumePro, Mitsubishi Electric's real-time volume rendering hardware for PCs. His research interests include computer graphics, scientific visualization, and computer architecture. Hanspeter Pfister received his Ph.D. in Computer Science in 1996 from the State University of New York at Stony Brook and his M.S. in Electrical Engineering from the Swiss Federal Institute of Technology (ETH) Zurich, Switzerland, in 1991. He is a member of the ACM, ACM SIGGRAPH, IEEE, the IEEE Computer Society, and the Eurographics Association.

Erik Piip
Manager, Computer Network Services, MERL Headquarters

Erik is the manager of the Computer Network Services Group. The group supports MERL's computing and network infrastructure and end-users. Erik is responsible for identifying MERL wide strategic and tactical enhancements. In past lives, Erik worked for Digital Equipment in multiple roles; service delivery, support, fault management and analysis, and server management design. Other interests include using and building large telescopes and instrumentation and Amateur Radio.

Fatih Porikli
Ph.D., Polytechnic University, 2001
Member Technical Staff, MERL Murray Hill

Fatih Porikli received the B.S. degree from the Bilkent University, Turkey in 1992, and the M.S. and Ph.D. degrees in electrical engineering from the Polytechnic University, Brooklyn, NY in 1995 and 2001, respectively. From 1999 to 2000, he worked for Hughes Research Labs, Malibu, CA. He joined to Mitsubishi Electric Research Labs at Murray Hill in August 2000. His research interests are in the areas of video processing, computer vision, 3-D depth estimation, aerial image processing, texture segmentation, robust optimization, network traffic, and digital signal filtering.

Stanley Pozerski
B.A. Computer Systems, Daniel Webster College 1987
Systems Network Administrator, MERL Headquarters

Stan's interests have followed the application of computers to a variety of manufacturing tasks including using PDP-11's to demonstrate control of multiple reactor chemical processes, using personal computers for production testing, manufacturing chemicals and controlling multi-axis rotary assembly machines. More recently, Stan has been Systems Administrator of a CIM system for a semiconductor facility performing shop floor scheduling, data collection, and process monitoring. Currently, Stan supports Windows and Linux clients and servers, networking, and the wide variety of PC applications used at MERL Cambridge Research.

Bhiksha Raj
Ph.D., Carnegie Mellon University, 2000
Research Scientist, MERL Cambridge Research

Dr. Bhiksha Raj joined CRL as a Staff Scientist. He completed his Ph.D. from Carnegie Mellon University in May 2000. Dr. Raj works mainly on algorithmic aspects of speech recognition, with special emphasis on improving the robustness of speech recognition systems to environmental noise. His latest work is on the use of statistical information about speech for the automatic design of filter-and-sum microphone arrays. Dr. Raj has over thirty conference and journal publications and is currently in the process of publishing a book on missing-feature methods for noise-robust speech recognition.

Ramesh Raskar
University of North Carolina at Chapel Hill
Research Scientist, MERL Cambridge Research

Ramesh Raskar joined MERL Cambridge Research as a Research Scientist in 2000. Prior to that, he was at the Office of the Future group at UNC's Computer Graphics lab. As part of his dissertation, he developed a framework for projector based 3D graphics by treating a projector as the dual of a camera. Current work includes topics from non-photorealistic rendering, computer vision and intelligent user interfaces. He is a member of the ACM and IEEE.

Charles Rich
Ph.D., Massachusetts Institute of Technology, 1980
Distinguished Research Scientist, MERL Cambridge Research

The thread connecting all of Dr. Rich's research has been to make interacting with a computer more like interacting with a person. As a founder and director of the Programmer's Apprentice project at the MIT Artificial Intelligence Laboratory in the 1980s, he pioneered research on intelligent assistants for software engineers. Dr. Rich joined MERL Cambridge Research in 1991 as a founding member of the research staff. For the past several years, he has been working on a technology for building collaborative interface agents, called COLLAGEN, which is based on a theory of human collaborative discourse. Dr. Rich is a Fellow and past Councilor of the American Association for Artificial Intelligence and was co-chair of AAAI'98.

David Rudolph
M.S., University of Illinois, 1989
Principal Technical Staff, MERL Cambridge Systems

David Rudolph has been at CSL for 11 years. During this time he has contributed to several systems software projects, including the gm80 simulator. His last 6 years have been spent developing the "Network Replication" project, which is now being successfully marketed by Veritas Software Corporation as VVR (Veritas Volume Manager). This project was the 1998 winner of the Corporate R&D GM's Award for Excellence. Before joining CSL, David spent three years at Data General, interrupted by a two-year stint at the University of Illinois, where his research interests were system software and performance analysis for massively parallel architectures.

Kathy Ryall
Ph.D., Harvard University, 1997
Principal Technical Staff, MERL Cambridge Systems

Kathy Ryall's research interests focus on user interfaces and improving human-computer collaboration. She is particularly interested in the design and implementation of systems in which the interface acts as a medium for people and computers to work together on solving problems, rather than as a means for people to control computers. Prior to joining MERL Cambridge Systems, Kathy served as an Assistant Professor of Computer Science at the University of Virginia for three years. While at MERL she has worked on a number of collaborative interaction technologies. She is currently working on DiamondTouch, a multi-user, multi-touch technology.

Zafer Sahinoglu
PhD, New Jersey Institute of Technology, 2001
Technical Staff, MERL Murray Hill

He received his B.S. in E.E. from Gazi Uni., Ankara, M.S. in BME and Ph.D. in EE from NJIT. He was awarded the Hashimoto Prize in 2002 by NJIT. He worked at AT&T Shannon Labs in 1999. He joined MERL in March 2001. He is currently with the VisComm Group at MHL. His research includes home networking, QoS issues and resource management in FGS video streaming, DSL technologies, traffic self-similarity and biomedical signal processing. He has published numerous conference papers in the relevant topics. He holds one European patent, and has filed 9 patents. Two journal submissions are under review.

Bent Schmidt-Nielsen
B.S. University of California at San Diego, 1971
Senior Principal Technical Staff, MERL Cambridge Systems

Bent Schmidt-Nielsen has seven years of experience at Dragon Systems in applying speech recognition to useful products. Here at MERL he is paying a lot of attention to making speech interfaces robust and usable. Bent has very broad interests in science and technology. Among many other activities he has taught genetics at the University of Massachusetts at Boston and he has been a leader in the development of an easy to use mass market database.

Derek Schwenke
M.S., Worcester Polytechnic Institute, 1988
Principal Technical Staff, MERL Cambridge Systems

Derek Schwenke received his B.S.E.E. from Tulane and M.S.C.S. from Worcester Polytechnic Institute. At Raytheon (1984-1988) in Marlboro, MA. Derek worked on image processing and satellite communications systems. At MERL Cambridge Systems (1988) Derek worked on the design and simulation of the M80 and PXB1 CPUs, and software development using the OSF-DCE/Encina platform. He co-developed the OpenMQ™ message queuing system for MELCO. Derek has worked on the Spline™ and OpenCommunity™ virtual reality system, co-developing the Schmoozer™ Java UI and the Internet Sharing and Transfer Protocol (ISTP™) extension. He's worked on the Concordia™ mobile agent project's Java security architecture and the Location Aware Systems project. Currently he is working on FormsTalk™ and Collagen multimodal interfaces and is a member of the W3C VoiceXML and Multimodal working groups.

Huairong Shao
Ph.D., Tsinghua University, 1999
Research Scientist, MERL Cambridge Research

Huairong Shao has research interests in adaptive and reliable multimedia communications, QoS provision for the next generation wired and wireless Internet, pervasive computing and collaborative systems. Before joining MERL Huairong Shao worked with Microsoft Research Beijing and Redmond. He received his Ph.D. in Computer Science from Tsinghua University in 1999.

Chia Shen
Ph.D., University of Massachusetts, 1992
Associate Director & Senior Research Scientist, MERL Cambridge Research

Dr Shen's research spans from non-traditional off-the-desktop interactive user interfaces and HCI for multi-user applications, to distributed real-time systems and multimedia systems in wired and wireless networks. Her long term research interest is to facilitate and enrich our communication from both the network level and the HCI perspective. Her most recent research projects include DiamondSpin, Personal Digital Historian(PDH) and MidART.

Samuel Shipman
M.Sc., Carnegie Mellon University, 1985
Principal Technical Staff, MERL Cambridge Systems

Sam Shipman received the M.S. degree in Computer Science from Carnegie Mellon University and the B.S. from UNC-Wilmington. His technical interests and background are in real-time and distributed operating systems research and development. At MERL Cambridge Systems, he has worked on the Network Replication and Open Community projects, and on smaller efforts related to MPEG-7, interactive surroundings, and fingerprint recognition.

Candace Sidner
Ph.D., Massachusetts Institute of Technology, 1979
Senior Research Scientist, MERL Cambridge Research

Candy Sidner is an expert in user interfaces, especially those involving speech and natural language understanding, and human and machine collaboration. Before coming to MERL, she had been a research staff member at Bolt Beranek Newman, Digital Equipment Corp., and Lotus Development Corp., and a visiting scientist at Harvard University. She is currently working on applying speech understanding technology to collaborative interface agents in the COLLAGEN project. Dr. Sidner was Chair of the 2001 International Conference on Intelligent User Interfaces and is a past President of the Association for Computational Linguistics. She is also a Fellow and past Councilor of the American Association for Artificial Intelligence.

Huifang Sun
Ph.D., University of Ottawa, 1986
Deputy Director, MERL Murray Hill

Huifang Sun received his B.Sc. degree from Harbin Engineering Institute, Harbin, China in 1967, and his Ph.D. from University of Ottawa, Canada, in 1986. He was an Associate Professor at Fairleigh Dickinson University before moving to Sarnoff Research Laboratory in 1990, where he was Technology Leader of Digital Video Communication. He joined MERL Murray Hill in 1995. His research interests include digital video/image compression and digital communication.

Jay Thornton
Ph.D., University of Michigan, 1982
Group Manager, Computer Vision Applications, MERL Cambridge Systems

Jay Thornton's degree program was Mathematical Psychology. His doctoral work focused on perception and vision, and his thesis concerned channels mediating color vision. After a post doc at the University of Pennsylvania, he worked for Polaroid Corporation, first in the Vision Research Laboratory and then as manager of the Image Science Laboratory. At Polaroid he worked on problems in color reproduction, image quality, image processing, and half toning. At MERL since January 2002, he manages the Computer Human Observation project, and is excited about the computer vision problems that arise when computers analyze, measure, count, detect, and recognize people.

Jeroen van Baar
M.Sc., Delft University of Technology, 1998
Research Associate, MERL Cambridge Research

Jeroen van Baar joined MERL Cambridge Research in 1999 as Member of Technical Staff. His interests are in the broad fields of computer graphics, scientific visualization and user interfaces. Jeroen is currently working on projector related work, combining computer vision and computer graphics. He has been a teaching assistant for (Introduction & Advanced) Computer Graphics for the past three years at Harvard Extension school.

Giovanni Vannucci
Ph.D., Columbia University, 1979
Senior Principal Technical Staff, MERL Murray Hill

Giovanni Vannucci receoved a Ph.D. in Electrical Engineering from Columbia University in 1979. He joined Bell Telephone Laboratories (then part of AT&T) and remained there for more than two decades that saw two AT&T divestitures and dramatic changes in communications technology. During that time he did research in microwave and satellite communications, optical communications and, more recently, in the area of wireless/portable communications and radiolocation. In 2000, Bell Laboratories opened a research lab in Italy and appointed Dr. Vannucci as Managing Director, which he held until his departure. Since October 2001 he has been at the Murray Hill Laboratory doing research in wireless networking, location techniques and Space-Time techniques for mobile communications.

Anthony Vetro
Ph.D., Polytechnic University, 2001
Senior Principal Technical Staff, MERL Murray Hill

Anthony Vetro received the B.S., M.S. and Ph.D. degrees in Electrical Engineering from Polytechnic University, Brooklyn, NY. He joined MERL in 1996, where he is currently a Senior Principal Member of the Technical Staff. At MERL, he worked on algorithms for down-conversion decoding of compressed video signals, which were implemented into MELCO's second-generation HDTV receiver chip. His current research interests are in the area of multimedia coding and transmission, with emphasis on content scaling and optimal resource allocation. He has also been an active participant in MPEG standards for several years.

Paul Viola
Ph.D., Massachusetts Institute of Technology, 1995
Research Scientist, MERL Cambridge Research

Before moving to MERL Paul Viola was an Associate Professor of Computer Science and Engineering at the Massachusetts Institute of Technology. He also spent two years as a visiting scientist in the Computational Neurobiology of the Salk Institute in San Diego. Paul has a broad background in advanced computational techniques, publishing in the fields of computer vision, neurobiological vision, medical imaging, mobile robotics, machine learning, and automated drug design. Paul was a recipient of a National Science Foundation Career award in 1998. He has worked on research and development with a number of companies including: Compaq, IBM Research, Arris Pharmaceuticals and Intarka.

Joseph Woelfel
M.S., Rutgers University, 1992
Principal Technical Staff, MERL Cambridge Systems

Before joining MERL in February 2001, Joe worked at Dragon Systems, where he led small teams developing an extensible voice architecture. In the years before that, Joe worked on the development of a statistical process control software package at GE-Fanuc. Joe earned a B.S. in Physics from SUNY Albany, and an M.S. in Communication and Information Science from Rutgers University. As Project Engineer for the Surveillance product, he will work on MERL's Interactive Surroundings Initiative.

Peter Wolf
B.S., Yale University, 1983
Senior Principal Technical Staff, MERL Cambridge Systems

Peter is an expert in Speech Technologies and a broad range of Software Engineering tools and practices. While Peter's role is often that of a technical expert and principal engineer, his main interest is the definition and creation of new products and services, made possible by new technologies. Peter is currently exploring the use of speech recognition to retrieve information with applications for cellphones, PDAs, automobiles and home entertainment.

David Wong
Ph.D., University of Connecticut, 1991
Group Manager, E-Services, MERL Cambridge Systems

David Wong is group manager of E-Services. This group consists of projects concerned with developing next generation agent technology (both intelligent and mobile) and data analysis and visualization tools. David's previous roles at MERL included serving as the lead architect and primary evangelist for Mitsubishi's flagship Concordia mobile agents framework. His technical interests include agent technology, data mining and visualization techniques, and distributed computing infrastructures. Prior to joining MERL in 1994, David worked on the advanced development and performance analysis of distributed transaction processing systems at Digital Equipment Corporation.

Christopher R. Wren
Ph.D., Massachusetts Institute of Technology, 2000
Research Scientist, MERL Cambridge Research

Christopher Wren's research area is Perception for Human-Computer Interaction. While Chris' recent work has focused on using computer vision techniques to create systems that are visually aware of the user, his current interests also extend to include audio processing and other sensing modalities. Prior to coming to MERL, he was at the Vision & Modeling Group at the MIT Media Laboratory. As part of his dissertation work, he developed a system for combining physical models with visual evidence in real time to recover subtle models of human motion.

Jonathan Yedidia
Ph.D., Princeton University, 1990
Research Scientist, MERL Cambridge Research

Jonathan Yedidia's graduate work at Princeton (1985-1990) and post-doctoral work at Harvard's Society of Fellows (1990-1993) focused on theoretical condensed-matter physics, particularly the statistical mechanics of systems with quenched disorder. From 1993 to 1997, he was a professional chess player and teacher. He then worked at the internet startup company Viaweb, where he helped develop the shopping search engine that has since become Yahoo's shopping service. In 1998, Dr. Yedidia joined MERL Cambridge Research Laboratory. He is particularly interested in the development of new methods to analyze graphical models. His work has applications in the fields of artificial intelligence, digital communications, and statistical physics.

William Yerazunis
Ph.D., Rensselaer Polytechnic Institute, 1987
Research Scientist, MERL Cambridge Research

William Yerazunis has worked in a number of fields including: optics and signal processing (for General Electric's jet engine manufacturing); computer graphics (at Rensselaer's Center for Interactive Computer Graphics); artificial intelligence and parallel symbolic computation (for DEC's OPS5, XCON, and the successor products such as RuleWorks); radioastronomy and exobiology (at Harvard University), and transplant immunology (for the American Red Cross). He holds 16 U.S. patents.

Fangfang Zhang
M.S., Brandeis University, 2000
Member Technical Staff, MERL Cambridge Systems

Fangfang Zhang received her Master in Software Engineering from Brandeis University. She joined CSL in January, 2001. She is currently working on product building and testing for Concordia development team. Prior to joining CSL, Fangfang worked shortly at CMGI, as a member of web-dialup service development team. She also worked at Scriptgen Pharmaceuticals, Inc., where she implemented a number of software and designed relational database to assist biophysics research.

Jinyun Zhang
Ph.D., University of Ottawa, 1991
Senior Principal Member Technical Staff, MERL Murray Hill

Jinyun received her Ph.D. in Electrical Engineering from the University of Ottawa in the area of digital signal processing, where she was also Visiting Scholar and worked on digital image processing and she was a teacher/lecturer at Tsinghua University, Beijing, China. Jinyun worked for Nortel Networks for ten years where she held engineering and management positions in the areas of VLSI design, Advanced Wireless Technology Development, Wireless Networks and Optical Networks. She has a broad technical background, specializing in system design, DSP algorithms, and real-time embedded S/W for wireless communications and DWDM optical networks. Jinyun joined the Murray Hill Laboratory in 2001.

Remo Ziegler
M.Sc., Swiss Federal Institute of Technology ETH, 2001
Member Technical Staff, MERL Cambridge Research

During Remo's studies of Computer Science, he specialized in computer graphics and computer vision. In term papers and his master thesis, he applied the knowledge in biomedical problems or virtual surgery. He gained most of his experience in computer graphics while preparing his master thesis, where he worked on a real time simulation of natural forces of skin and tissue as they resist incision by a scalpel. Remo joined MERL Cambridge Research in 2001 and is currently working on a program generating a 3D polygonal model out of several projections.

Recent Major Publications

The following lists the major publications by members of the MERL staff over the past 2 years. A publication is considered major if it appeared in a refereed journal or a refereed conference proceedings. For completeness, the list includes a number of publications that have been accepted for publication in the near future.

2002

Zhang, X.M.; Vetro, A.; Shi, Y.Q.; Sun, H., "Constant Quality Constrained Rate Allocation for FGS Coded Video," *IEEE Transactions on Circuits and Systems for Video Technology*, To Appear Fall 2002, Vol 12.

Zwicker, M.; Pfister, H.; van Baar, J.; Gross, M., "EWA Splatting," *IEEE Transactions on Visualization and Computer Graphics*, To Appear Fall 2002, Vol 12.

Shen, C.; Lesh, N.B.; Vernier, F.; Forlines, C.; Frost, J., "Sharing and Building Digital Group Histories," *ACM Conference on Computer Supported Cooperative Work (CSCW)*, To Appear November 2002, (TR2002-07).

Frisken, S.F.; Perry, R.N., "Efficient Estimation of 3D Euclidean Distance Fields from 2D Range Images," *IEEE/SIGGRAPH Symposium on Volume Visualization and Graphics (VolVis)*, To Appear October 2002.

Guo, J.; Matsubara, F.; Cukier, J.; Kong, H., "DTV for Personalized Mobile Access and Unified Home Control," *Prof Personal Wireless Communications*, To Appear October 2002.

Wong, D., "Improving Circuit Breaker Maintenance Management Tasks by Applying Mobile Agent Software Technology," *IEEE/PES Transmission and Distribution Conference and Exhibition (PES T&D)*, To Appear October 2002.

Raj, B.; Singh, R., "Classifier-Based Non-Linear Projection for Adaptive Endpointing of Continuous Speech," *Computer Speech and Language*, To Appear September 2002.

Bandyopadhyay, D.; Raskar, R.; Fuchs, H., "Dynamic Shader Lamps: Painting on Movable Objects," *IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, To Appear September 2002, (TR2001-50)

Horng, J.H.; Vannucci, G.; Zhang, J., "Capacity Enhancement for HSDPA in W-CDMA System," *IEEE Vehicular Technology Conference (VTC)*, To Appear September 2002.

Sun, X.; Manjunath, B.S.; Divakaran, A., "Representation of Motion Activity in Hierarchical Levels for Video Indexing and Filtering," *IEEE International Conference on Image Processing (ICIP)*, To Appear September 2002.

Divakaran, A.; Radhakrishnan, R.; Peker, K.A., "Motion Activity Based Extraction of Key Frames from Video Shots," *IEEE International Conference on Image Processing (ICIP)*, To Appear September 2002.

Lee, J.W.; Vetro, A.; Wang, Y.; Ho, Y.S., "MPEG-4 Video Object-Based Rate Allocation with Variable Temporal Rates," *IEEE International Conference on Image Processing (ICIP)*, To Appear September 2002.

Vetro, A.; Chen, C.W., "Rate-Reduction Transcoding Design for Wireless Video Streaming," *IEEE International Conference on Image Processing (ICIP)*, To Appear September 2002.

Porikli, F.M., "Automatic Threshold Determination of Centroid-Linkage Region Growing by MPEG-7 Dominant Color Descriptors," *IEEE International Conference on Image Processing (ICIP)*, To Appear September 2002.

Sidner, C.L.; Forlines, C., "Subset Languages for Conversing with Collaborative Interface Agents," *International Conference on Spoken Language Processing (ICSLP)*, To Appear September 2002.

Zhang, J.; Li, L.; Poon, T.C., "Turbo Coded HSDPA Systems with Transmit Diversity over Frequency Selective Fading Channels," *Personal Indoor Mobile Radio Communications (PIMRC)*, To Appear September 2002.

Divakaran, A., "Video Summarization and Indexing Using Combinations of the MPEG-7 Motion Activity Descriptor and Other Audio-Visual Descriptors," *Thyrrhenian International Workshop on Digital Communications (IWDC)*, To Appear September 2002.

Ren, L.; Pfister, H., Zwicker, M., "Object Space EWA Splatting: A Hardware-Accelerated Approach to High-Quality Point Rendering," *Eurographics*, To Appear September 2002.

Karpenko, O.; Hughes, J.F.; Raskar, R., "Free-Form Sketching with Variational Implicit Surfaces," *Eurographics (EG)*, To Appear September 2002, (TR2002-27).

Porikli, F.; Wang, Y., "Constrained Region Extraction of Video Objects by Color Masks and MPEG-7 Descriptors," *IEEE International Conference on Multimedia and Expo (ICME)*, To Appear August 2002.

Kong, H.S.; Guan, L.; Kung, S.Y., "A Self-Organizing Tree Map Approach for Image Segmentation," *International Conference on Signal Processing (ICSP)*, To Appear August 2002.

Moghaddam, B.; Tian, Q.; Lesh, N.B.; Shen, C.; Huang, T.S., "PDH: A Human-Centric Interface for Image Libraries," *IEEE International Conference on Multimedia and Expo (ICME)*, To Appear August 2002.

Shao, H.R.; Shen, C.; Kong, H.S., "Adaptive Pre-Stored Video Streaming with End System Multicast," *International Conference on Signal Processing (ICSP)*, To Appear August 2002.

Wolf, P.; Raj, B., "The MERL SpokenQuery Information Retrieval System: A System for Retrieving Pertinent Documents from a Spoken Query," *IEEE International Conference on Multimedia and Expo (ICME)*, To Appear August 2002.

Moghaddam, B.; Zhou, X., "Factorized Local Appearance Models," *IEEE International Conference on Pattern Recognition (ICPR)*, To Appear August 2002.

Gonzalez, O.; Ramamritham, K.; Shen, C.; Fohler, G., "Building Real-Time Systems with Commercial-Off-the-Shelf Components," *Building Reliable Component-Based Systems*, To Appear July 2002.

Peker, K.A.; Divakaran, A., "A Framework for Measurement of the Intensity of Motion Activity of Video Segments," *SPIE Photonics East Conference Internet Multimedia Management Systems III*, To Appear July 2002.

Garland, A.; Lesh, N.B., "Plan Evaluation with Incomplete Action Descriptions," *National Conference on Artificial Intelligence (AAAI)*, To Appear July 2002, (TR2002-05).

Klau, G.W.; Lesh, N.B.; Marks, J.W.; Mitzenmacher, M., "Human-Guided Tabu Search," *National Conference on Artificial Intelligence (AAAI)*, To Appear July 2002, (TR2002-09).

Matusik, W.; Pfister, H.; Ngan, A.; Ziegler, R.; McMillan, L., "Acquisition and Rendering of Transparent and Refractive Objects," *Eurographics Workshop on Rendering*, To Appear July 2002, (TR2002-30).

Matusik, W.; Pfister, H.; Beardsley, P.A.; Ngan, A.; Ziegler, R.; McMillan, L., "Image-Based 3D Photography Using Opacity Hulls," *ACM SIGGRAPH*, To Appear July 2002, (TR2002-29).

Radhakrishnan, R.; Divakaran, A., "Automatic Detection of Talking Head Segments in News Video in the Compressed Domain," *World Multi-Conference on Systemics, Cybernetics and Informatics (ISAS-SCI)*, To Appear July 2002.

Kuwahara, N.; Hata, T.; Vetro, A., "Map-Top Video Surveillance with Adaptive Video Delivery," *World Multi-Conference on Systemics, Cybernetics and Informatics (ISAS-SCI)*, To Appear July 2002.

You, C.; Gu, D.; Zhang, J., "Timing Synchronization for OFDM," *World Multi-Conference on Systemics, Cybernetics and Informatics (ISAS-SCI)*, To Appear July 2002.

Porikli, F.M.; Sahinoglu, Z., "An Online Renegotiation-Based Bandwidth Management with Circuit Assignment for VBR Traffic in Communication Networks," *World Multi-Conference on Systemics, Cybernetics and Informatics (ISAS-SCI)*, To Appear July 2002.

Kong, H.S.; Guan, L.; Kung, S.Y., "An Efficient Codebook Design Method of Vector Quantizer via Using Self-Organizing Tree Map," *World Multi-Conference on Systemics, Cybernetics and Informatics (ISAS-SCI)*, To Appear July 2002.

Li, L.; Zhang, J.; Horng, J.H., "Channel Estimation Error Impact on the Link Level Performance of Space-Time Coded Systems," *World Multi-Conference on Systemics, Cybernetics and Informatics (ISAS-SCI)*, To Appear July 2002.

Cao., L.; Orlik, P.V.; Gu, D.; Zhang, J., "A Trellis Based Technique for Blind Channel Estimation and Signal Detection," *World Multi-Conference on Systemics, Cybernetics and Informatics (ISAS-SCI)*, To Appear July 2002.

Orlik, P.V.; Derand, T.; Horng, J.H.; Wang, D., "Study of Rate 1/4 HSDPA Turbo Codes and Scaling Effect," *World Multi-Conference on Systemics, Cybernetics and Informatics (ISAS-SCI)*, To Appear July 2002.

Wong, D., "Agent-Oriented Approach to Work Order Management for Circuit Breaker Maintenance," *IEEE Power Engineering Society Summer Meeting (PES)*, To Appear July 2002.

Tian, Q.; Moghaddam, B.; Huang, T.S., "Visualization, Estimation and User-Modeling for Interactive Browsing of Image Libraries," *International Conference on Image and Video Retrieval (CIVR)*, To Appear July 2002, (TR2001-28).

Rich, C.; Lesh, N.B.; Rickel, J.; Garland, A., "A Plug-in Architecture for Generating Collaborative Agent Responses," *International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, To Appear July 2002, (TR2002-10).

Radhakrishnan, R.; Divakaran, A., "A Date Reduction Procedure for 'Principal Cast' and other 'Talking Head' Detection," *World Multi-Conference on Systemics, Cybernetics and Informatics (ISAS-SCI)*, To Appear July 2002.

Moghaddam, B., "Principal Manifolds and Probabilistic Subspaces for Visual Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, Vol 24, Issue 6, pps 780-788, June 2002, (TR2002-13).

Moghaddam, B.; Zhou, X., "Factorization for Probabilistic Local Appearance Models," *IASTED International Conference on Signal Processing, Pattern Recognition and Applications (SPPRA)*, June 2002.

Divakaran, A.; Cabasson, R., "Content-Based Browsing System for Personal Video Recorders," *IEEE International Conference on Consumer Electronics (ICCE)*, Digest of Technical Papers, pps 114-115, June 2002.

Vetro, A.; Hata, T.; Kuwahara, N.; Kalva, H., "Complexity-Quality Analysis of MPEG-2 to MPEG-4 Transcoding Architectures," *IEEE International Conference on Consumer Electronics (ICCE)*, Digest of Technical Papers, pps 130-131, June 2002.

Lin, Y-C.; Wang, C-N.; Chiang, T.; Vetro, A.; Sun, H., "Efficient FGS to Single Layer Transcoding," *IEEE International Conference on Consumer Electronics (ICCE)*, Digest of Technical Papers, pps 134-135, June 2002.

Horng, J.H.; Zhang, J., "Signal Circulation for Adaptive Equalization in Digital Communication Systems," *IEEE International Conference on Consumer Electronics (ICCE)*, Digest of Technical Papers, pps 300-301 June 2002.

Sahinoglu, Z.; Peker, K.A.; Matsubara, F.; Cukier, J., "A Mobile Network Architecture with Personalized Instant Information Access," *IEEE International Conference on Consumer Electronics (ICCE)*, Digest of Technical Papers, pps 34-35 June 2002.

Quigley, A.; Leigh, D.L.; Lesh, N.B.; Marks, J.W.; Ryall, K.; Wittenburg, K., "Semi-Automatic Antenna Design via Sampling and Visualization," *IEEE AP-S International Symposium and USNC/URSI National Radio Science Meeting (APS/URSI)*, June 2002.

Dietz, P.H.; Yerazunis, W.S.; Leigh, D.L., "Wireless Level-Sensing Restaurant Glassware," *IEEE Sensors*, June 2002, (TR2002-21).

Rickel, J.; Lesh, N.B.; Rich, C.; Sidner, C.L.; Gertner, A., "Collaborative Discourse Theory as a Foundation for Tutorial Dialogue," *International Conference on Intelligent Tutoring Systems (ITS)*, Vol 2363, pps 542-551, June 2002.

Raskar, R.; Ziegler, R.; Willwacher, T., "Cartoon Dioramas in Motion," *International Symposium on Non-Photorealistic Animation and Rendering (NPAR)*, June 2002, (TR2002-28).

Brand, M.E., "Subspace Mappings for Image Sequences," *Statistical Methods in Video Processing*, June 2002, (TR2002-25).

Singh, R.; Stern, R.M.; Raj, B., "Signal and Feature Compensation Methods for Robust Speech Recognition," *CRC Handbook on Noice Reduction in Speech Applications*, May 2002.

Singh, R.; Raj, B.; Stern, R.M., "Model Compensationa and Matched Condition Methods for Robust Speech Recognition," *CRC Handbook on Noise Reductionin Speech Applications*, May 2002.

Moghaddam, B.; Yang, M-H, "Learning Gender with Support Faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, Vol 24, Issue 5, pps 707-711, May 2002, (TR2002-12).

Gu, D.; Zhang, J.; You, C.; Miyake, M.; Ma, X., "An OFDM Receiver Structure for Broadband Wireless LANs," *International Conference on Third Generation Wireless and Beyond*, May 2002.

Ma, X.; Kobayashi, H.; Schwartz, S.C.; Zhang, J.; Gu, D., "Time Domain Channel Estimation for OFDM Using EM Algorithm," *International Conference on Third Generation Wireless and Beyond*, May 2002.

Kung, S.Y.; Zhang, X.; Vannucci, G.; Zhang, J., "Sequential Bezout Equalizer for MIMO Systems," *International Conference on Third Generation Wireless and Beyond*, May 2002.

Cao, L.; Zhang, J.; Chen, C.W,; Orlik, P.V.; Miyake, M., "Study of Two 1/4 Turbo Code Receivers in Non-Frequency Selective Fast Fading Channels," *International Conference on Third Generation Wireless and Beyond*, May 2002.

Brand, M.E., "Incremental Singular Value Decomposition of Uncertain Data with Missing Values," *European Conference on Computer Vision (ECCV)*, Vol 2350, pps 707-720, May 2002, (TR2002-24).

Vetro, A.; Yin, P.; Liu, B.; Sun, H., "Reduced Spatio-Temporal Transcoding Using an Intra Refresh Technique," *IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2002.

Zhang, X.M.; Shi, Y.Q.; Chen, H.; Hamovich, A.; Vetro, A.; Sun, H., "Successive Packing Based Interleaver for Turbo Codes," *IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2002.

Klau, G.W.; Lesh, N.B.; Marks, J.W.; Mitzenmacher, M.; Schafer, G.T., "The HuGS Platform: A Toolkit for Interactive Optimization," *Advanced Visual Interfaces (AVI)*, May 2002, (TR2002-08).

Vernier, F.; Lesh, N.B.; Shen, C., "Visualization Techniques for Circular Tabletop Interfaces," *ACM Advanced Visual Interfaces (AVI)*, May 2002, (TR2002-01).

Shakhnarovich, G.; Viola, P.; Moghaddam, B., "A Unified Learning Framework for Face Detection and Classification," *International Conference on Automatic Face and Gesture Recognition (FG)*, May 2002.

Xie, L.; Chang, S.; Divakaran, A.; Sun, H., "Structure Analysis of Soccer Video with Hidden Markov Models," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2002.

Vetro, A.; Wang, Y.; Sun, H., "A Probabilistic Approach for Rate-Distortion Modeling of Multi-scale Binary Shape," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2002.

Seltzer, M.; Raj, B.; Stern R., "Speech Recognizer Based Microphone Array Processing for Robust Hands-Free Speech Recognition," *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, May 2002.

Cao, L.; Chen, C.; Zhang, J.; Orlik, P.V.; Gu, D., "Blind Channel Estimation and Equalization Using Viterbi Algorithms," *IEEE Vehicular Technology Conference (VTC)*, Vol 3, pps 1532-1535, May 2002.

Scott, S.D.; Lesh, N.B.; Klau, G.W., "Investigating Human-Computer Optimization," *ACM Conference on Human Factors in Computing Systems (CHI)*, ISBN: 158113-453-3, pps 155-162, April 2002, (TR2001-39).

Sahinoglu, Z.; Tekinay, S., "Efficient Paramater Selection for Use of Self-Similarity in Real Time Resource Management," *IEEE International Conference on Communications (ICC)*, April 2002.

Terry, M.; Mynatt, E.D.; Ryall, K.; Leigh, D.L., "Social Net: Using Patterns of Physical Proximity Over Time to Infer Shared Interests," *ACM Conference on Human Factors in Computing Systems (CHI)*, ISBN: 1-58113-454-1, pps 816-817, April 2002.

Freeman, W.T.; Jones, T.R.; Pasztor, E.C., "Example-Based Super-Resolution," *IEEE Computer Graphics and Applications*, March/April 2002.

Wang, D., Orlik, P.V., Derand, T., Cao, L., Kobayashi, H.; Zhang, J., "Improved Low-Complexity Turbo Decoding for HSDPA Applications," *Annual Conference on Information Sciences and Systems (ISS)*, March 2002.

Vetro, A.; Devillers, S., "Position Paper: Delivery Context in MPEG-21," *W3C Delivery Context Workshop*, March 2002.

Yedidia, J.S., "Generalized Belief Propagation and Free Energy Minimization," *Information Theory Workshop at Mathematical Sciences Research Institute (MSRI)*, March 2002.

Divakaran, A.; Radhakrishnan, R., "Data Reduction Procedure for 'Principal-Cast' and other 'Talking-Head' Detection," *SPIE Conference on Storage and Retrieval for Media Databases*, Vol 4676, pps 177-182, January 2002.

Peker, K.A.; Cabasson, R.; Divakaran, A., "Rapid Generation of Sports Video Highlights Using the MPEG-7 Motion Activity Descriptor," *SPIE Conference on Storage and Retrieval for Media Databases*, Vol 4676, pps 318-323, January 2002.

Raskar, R.; van Baar, J.; Chai, J.X., "A Low-Cost Projector Mosaic with Fast Registration," *Asian Conference on Computer Vision (ACCV)*, January 2002, (TR2002-14).

Lee, J.W.; Vetro, A.; Wang, Y.; Ho, Y.S., "Object-Based Rate Allocation with Spatio-Temporal Trade-Offs," *SPIE Conference on Visual Communications and Image Processing (VCIP)*, Vol 4671, pps 374-384, January 2002.

Zhang, X.; Vetro, A.; Shi, Y.Q.; Sun, H., "Constant Quality Constrained Rate Allocation for FGS Video Coded Bitstreams," *SPIE Conference on Visual Communications and Image Processing (VCIP)*, Vol 4671, pps 817-827, January 2002.

Yin, P.; Vetro, A.; Sun, H.; Liu, B., "Drift Compensation Architectures and Techniques for Reduced Resolution Transcoding," *SPIE Conference on Visual Communications and Image Processing (VCIP)*, Vol 4671, pps 180-191, January 2002.

Bhatti, G.; Sahinoglu, Z.; Peker, K.A.; Guo, J.; Matsubara, F., "A TV-Centric Home Network to Provide a Unified Access to UPnP and PLC Domains," *IEEE International Workshop on Networked Appliances (IWNA)*, January 2002.

Eisenstein, J.; Rich, C., "Agents and GUI's from Task Models," *ACM International Conference on Intelligent User Interfaces*, ISBN: 1-58113-382-0 pps 47-54, January 2002, (TR2002-16).

Sidner, C.L.; Dzikovska, M., "Hosting Activities: Experience with and Future Directions for a Robot Agent Host," *ACM International Conference on Intelligent User Interfaces*, ISBN: 1-58113-382-0, pps 143-150, January 2002, (TR2002-03).

2001

Casey, M.A., "General Sound Similarity and Sound Recognition Tools," *Introduction to MPEG-7: Multimedia Content Description Language*, 2001.

Freeman, W.T.; Haddon, J.A.; Pasztor, E.C., "Learning Motion Analysis," *Statistical Theories of the Brain*, edited by R. Rao, B. Olshausen and M. Lewicki, 2001, (TR2000-32).

Rich, C.; Sidner, C.L.; Lesh, N.B., "COLLAGEN: Applying Collaborative Discourse Theory to Human-Computer Interaction," *Artificial Intelligence Magazine*, Winter 2001, (TR2000-38).

Raj, B.; Migdal, J.; Singh, R., "Distributed Speech Recognition with Codec Parameters," *IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, December 2001, (TR2001-45).

Brand, M.E., "Morphable 3D Models from Video," *IEEE Computer Vision and Pattern Recognition (CVPR)*, December 2001, (TR2001-37).

Brand, M.E.; Bhotika, R., "Flexible Flow for 3D Nonrigid Tracking and Shape Recovery," *IEEE Computer Vision and Pattern Recognition (CVPR)*, December 2001, (TR2001-38).

Raskar, R.; Beardsley, P., "A Self-Correcting Projector," *IEEE Computer Vision and Pattern Recognition (CVPR)*, December 2001, (TR2001-46).

Porikli, F.M., "Image Simplification by Robust Estimator Based Reconstruction Filter," *International Symposium on Computer and Information Sciences (ISCIS)*, November 2001.

Porikli, F.M., "Video Object Segmentation by Volume Growing Using Feature-Based Motion Estimator," *International Symposium on Computer and Information Sciences (ISCIS)*, November 2001.

Raskar, R.; Low, K-L, "Interacting with Spatially Augmented Reality," *ACM International Conference on Virtual Reality, Computer Graphics and Visualization in Africa (AFRIGRAPH)*, November 2001, (TR2001-51).

Sahinoglu, Z., "Novel Adaptive Bandwidth Allocation Algorithm: Wavelet-Decomposed Signal Energy Approach," *IEEE Global Telecommunications Conference (Globecom)*, November 2001.

Yerazunis, W.S.; Carbone, M., "Privacy-Enhanced Displays by Time-Masking Images," *Australian Conference on Computer-Human Interaction (OzCHI)*, November 2001, (TR2002-11).

Dietz, P.H.; Leigh, D.L., "DiamondTouch: A Multi-User Touch Technology," *ACM Symposium on User Interface Software and Technology (UIST)*, ISBN: 1-58113-438-X, pps 219-226, November 2001.

Dietz, P.H.; Yerazunis, W.S., "Real-Time Audio Buffering for Telephone Applications," *ACM Symposium on User Interface Software and Technology (UIST)*, ISBN: 1-58113-438-X, pps 193-194, November 2001.

Divakaran, A.; Regunathan, R.; Peker, K.A., "Video Summarization Using Descriptors of Motion Activity: A Motion Activity Based Approach to Key-Frame Extraction from Video Shots," *Journal of Electronic Imaging*, October 2001.

Sun, X.; Divakaran, A.; Manjunath, B.S., "A Motion Activity Descriptor and Its Extraction in the Compressed Domain," *IEEE Pacific-Rim Conference on Multimedia (PCM)*, LNCS 2195, pps 450, October 2001.

Wu, Y.; Vetro, A.; Sun, H.; Kung, S.Y., "Intelligent Multi-Hop Video Communications," *IEEE Pacific-Rim Conference On Multimedia (PCM)* , LNCS 2195, pps 716, October 2001.

Garland, A.; Ryall, K.; Rich, C., "Learning Hierarchical Task Models by Defining and Refining Examples," *ACM International Conference on Knowledge Capture (KCAP)*, ISBN: 1-58113-380-4, pps 44-51, October 2001, (TR2001-26).

Zwicker, M.; Pfister, H.; van Baar, J.; Gross, M., "EWA Volume Splatting," *IEEE Visualization (VIS)*, October 2001, (TR2001-27).

Porikli, F.M.; Sahinoglu, Z., "Dynamic Bandwidth Allocation with Optimal Number of Renegotiations in ATM Networks," *International Conference on Computer Communications and Networks (ICCCN)*, October 2001.

Vetro, A.; Wang, Y.; Sun, H., "Rate-Distortion Optimized Video Coding Considering Frameskip," *IEEE International Conference on Image Processing (ICIP)*, October 2001.

Porikli, F.M., "Accurate Detection of Edge Orientation for Color and Multi-Spectral Imagery," *IEEE International Conference on Image Processing (ICIP)*, October 2001.

Porikli, F.M.; Wang, Y., "An Unsupervised Multi-Resolution Object Extraction Algorithm Using Video-Cubes," *IEEE International Conference on Image Processing (ICIP)*, October 2001.

Horng, J.H.; Nakache, Y-P., "Adaptive Diversity Combining for Interference Cancellation in CDMA Systems," *IEEE Vehicular Technology Conference (VTC)*, Vol 3, pps 1899-1902, October 2001.

Moghaddam, B.; Zhou, X.; Huang, T.S., "ICA-Based Probabilistic Local Appearance Models," *IEEE International Conference on Image Processing (ICIP)*, October 2001, (TR2001-29).

Peker, K.A.; Divakaran, A.; Sun, H., "Constant Pace Skimming and Temporal Sub-sampling of Video Using Motion Activity," *IEEE International Conference on Image Processing (ICIP)*, October 2001.

Yedidia, J.S.; Sudderth, E.B.; Bouchaud, J-P., "Projection Algebra Analysis of Error-Correcting Codes," *Allerton Conference on Communication, Control, and Computing*, pps 662-671, October, 2001, (TR2001-35).

Moghaddam, B.; Tian, Q.; Lesh, N.B.; Shen, C.; Huang, T.S., "Visualization and Layout for Personal Photo Libraries," *International Workshop on Content-Based Multimedia Indexing (CBMI)*, September 2001.

Wang, D.; Kobayashi, H.; Zhang, J.; Gu, D., "On Bi-Directional SOVA Decoding of Turbo Codes," *IEICE Conference*, September 2001.

Tian, Q.; Moghaddam, B.; Huang, T.S., "Display Optimization for Image Browsing," *International Workshop Multimedia Databases and Image Communication*, Vol 2184, pps 167-178, September 2001.

Divakaran, A., "An Overview of MPEG-7 Motion Descriptors and Their Applications," *International Conference on Computer Analysis of Images and Patterns (CAIP)*, LNCS 2124, pps 29, September 2001.

Porikli, F.M., "Object Segmentation of Color Video Sequences," *International Conference on Computer Analysis of Images and Pattern (CAIP)* , LNCS 2124, pps 610, September 2001.

Raj, B.; Seltzer, M.; Stern, R., "Robust Speech Recognition: The Case for Restoring Missing Features," *Workshop for Consistent and Reliable Acoustic Cues (CRAC)*, September 2001, (TR2001-40).

Casey, M.A., "Reduced-Rank Spectra and Entropic Priors as Consistent and Reliable Cues for General Sound Recognition," *Workshop for Consistent & Reliable Acoustic Cues (CRAC)*, September 2001.

Casey, M.A., "General Sound Similarity and Recognition: Musical Applications of MPEG-7 Audio," *Organized Sound*, August 2001.

Xu, P.; Chang, S.F.; Divakaran, A.; Vetro, A.; Sun, H., "Algorithm and System for High-Level Structure Analysis and Event Detection in Soccer Video," *IEEE International Conference on Multimedia and Expo (ICME)*, August 2001.

Vetro, A.; Wang, Y.; Sun, H., "Estimating Distortion for Coded and Non-Coded Frames for Frameskip-Optimized Video Coding," *IEEE International Conference on Multimedia and Expo (ICME)*, August 2001.

Peker, K.A.; Divakaran, A., "A Novel Pair-Wise Comparison Based Analytical Framework for Automatic Measurement of Intensity of Motion Activity of Video," *IEEE International Conference on Multimedia and Expo (ICME)*, August 2001.

Efros, A.A.; Freeman, W.T., "Image Quilting for Texture Synthesis and Transfer," *ACM SIGGRAPH*, ISBN: 1-58113-374-X, pps 341-346, August 2001, (TR2001-17).

Frisken, S.F.; Perry, R.N., "A Computationally Efficient Framework for Modeling Soft Body Impact," *ACM SIGGRAPH*, August 2001, (TR2001-11).

Pope, J.; Frisken, S.F.; Perry, R.N., "Dynamic Meshing Using Adaptively Sampled Distance Fields," *ACM SIGGRAPH*, August 2001, (TR2001-13).

Frisken, S.F.; Perry, R.N., "Computing 3D Geometry Directly from Range Images," *ACM SIGGRAPH*, August 2001, (TR2001-10).

Perry, R.N., "A New Method for Numerical Constrained Optimization," *ACM SIGGRAPH*, August 2001, (TR2001-14).

Jones, T.R.; Perry, R.N.; Callahan, M., "Shadermaps: A Method for Accelerating Procedural Shading," *ACM SIGGRAPH*, August 2001, (TR2000-25).

Perry, R.N.; Frisken, S.F., "Kizamu: A System for Sculpting Digital Characters," *ACM SIGGRAPH*, ISBN: 1-58113-374-X, pps 47-56, August 2001, (TR2001-08).

Zwicker, M.; Pfister, H.; van Baar, J.; Gross, M., "Surface Splatting," *ACM SIGGRAPH*, ISBN: 1-58113-374-X, pps 371-378, August 2001, (TR2001-20).

Raskar, R., "Hardware Support for Non-Photorealistic Rendering: i.3.5," *SIGGRAPH/Eurographics Workshop on Graphics Hardware (HWWS)*, ISBN: 1-58113-407-X, pps 410-46i.3, August 2001, (TR2001-23).

Rich, C.; Lesh, N.B.; Sidner, C.L., "Human-Computer Collaboration for Universal Access," *International Conference on Universal Access in Human-Computer Interaction (UAHCI)*, Invited Paper, August 2001.

Yedidia, J.S.; Freeman, W.T.; Weiss, Y., "Understanding Belief Propagation and Its Generalizations," *International Joint Conference on Artificial Intelligence (IJCAI)*, Distinguished Presentations, August 2001, (TR2001-22).

Chen, P.; Horng, J.H.; Bao, J., Kobayashi, H., "A Joint Channel Estimation and Signal Detection Algorithm for OFDM System," *International Symposium on Signals, Systems and Electronics (ISSSE)*, July 2001.

Horng, J.; Nakache, Y-P., "Using Diversity Combining for Adaptive Interference Cancellation in CDMA Systems," *International Symposium on Signals, Systems and Electronics (ISSSE)*, July 2001.

Lesh, N.B.; Rich, C.; Sidner, C.L., "Collaborating with Focused and Unfocused Users Under Imperfect Communication," *International Conference on User Modelling* , July 2001, (TR2000-39).

Bell, M.; Freeman, W.T., "Learning Local Evidence for Shading and Reflectance," *IEEE International Conference on Computer Vision (ICCV)*, Vol 1, pps 670-677, July 2001, (TR2001-04).

Cassell, J.; Nakano, Y.I.; Bickmore, T.W.; Sidner, C.L.; Rich, C., "Non-Verbal Cues for Discourse Structure," *Association for Computational Linguistics Annual Conference (ACL)*, pps 106-115, July 2001, (TR2001-24).

Jeannin, S.; Divakaran, A., "MPEG-7 Visual Motion Descriptors," *IEEE Transactions on Circuits and Systems for Video Technology*, June 2001.

Casey, M.A., "MPEG-7 Sound-Recognition Tools," *IEEE Transactions on Circuits and Systems for Video Technology*, June 2001.

Moghaddam, B.; Biermann, H.; Margaritis, D., "Regions-of-Interest and Spatial Layout for Content-Based Image Retrieval," *Multimedia Tools and Applications*, Vol 14, No 2, pps 201-210, June 2001, (TR2000-35).

Zhang, X.; Vetro, A.; Sun, H.; Shi, Y-Q., "Robust Decoding for Reduced Error Propagation of DC/AC Prediction Errors," *Workshop and Exhibition on MPEG-4 (WEMP4)* , June 2001.

Raskar, R.; Welch, G.; Low, K-L.; Bandyopadhyay, D., "Shader Lamps: Animating Real Objects with Image Based Illumination," *Eurographics Workshop on Rendering*, June 2001, (TR2001-21).

Vetro, A.; Divakaran, A.; Sun, H., "Providing Multimedia Services to a Diverse Set of Consumer Devices," *IEEE International Conference on Consumer Electronics (ICCE)*, pps 32-33, June 2001.

Horng, J.H.; Derand, T., "Interference Cancellation in CDMA Downlink Supporting Variable Rate Services," *IEEE International Conference on Consumer Electronics (ICCE)*, pps 100-101, June 2001.

You, C.; Horng, J.H., "Optimum Frame and Frequency Synchronization for OFDM Systems," *IEEE International Conference on Consumer Electronics (ICCE)*, pps 226-227, June 2001.

Wang, H.; Shen, C.; Shin, K., "Adaptive-Weighted Packet Scheduling for Premium Service," *IEEE International Conference on Communications (ICC)*, Vol 6, pps 1846-1850, June 2001, (TR2001-07).

Moghaddam, B.; Nastar, C.; Pentland, A., "A Bayesian Similarity Measure for Deformable Image Matching," *Image and Vision Computing*, Vol. 19, Issue 5, pps 235-244, May 2001, (TR2001-52).

Yedidia, J.S.; Bouchaud, J-P., "Renormalization Group Approach to Error-Correcting Codes," *Workshop on Statistical Physics and Capacity-Approaching Codes*, May 2001, (TR2001-19).

Rickel, J.; Lesh, N.B.; Rich, C.; Sidner; C.L.; Gertner, A., "Building a Bridge Between Intelligent Tutoring and Collaborative Dialogue Systems," *International Conference on Artificial Intelligence in Education (AI-ED)*, Outstanding Paper Award, pps 592-594, May 2001, (TR2000-40).

Vetro, A.; Sun, H., "Media Conversions to Support Mobile Users," *IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, May 2001.

Vetro, A.; Sun, H., "Encoding and Transcoding Multiple Video Objects with Variable Temporal Resolution," *IEEE International Symposium on Circuits and Systems (ISCAS)*, Vol 5, pps 21-24, May 2001.

Kuhn, P.; Suzuki, T.; Vetro, A., "MPEG-7 Transcoding Hints for Reduced Complexity and Imrpoved Quality," *International Packet Video Workshop*, April 2001.

Freeman, W.T., "Review of Biometrics: Personal Identification in a Networked Society," *Pattern Analysis and Applications*, Vol 4, Issue 1, pps 67-68, April 2001, (TR2000-07).

Vetro, A.; Sun, H., "An Overview of MPEG-4 Object-Based Encoding Algorithms," *International Conference on Information Technology: Coding and Computing*, April 2001.

Vetro, A., Sun, H.; Wang, Y., "Object-Based Transcoding for Adaptable Video Content Delivery," *IEEE Transactions on Circuits and Systems for Video Technology*, March 2001.

Shen, C.; Lesh, N.B.; Moghaddam, B.; Beardsley, P.A.; Bardsley, R.S., "Personal Digital Historian: User Interface Design," *ACM Conference on Human Factors in Computing Systems (CHI)*, March 2001, (TR2000-34).

Aish, R.; Frankel, J.L.; Frazer, J.H.; Patera, A.T.; Marks, J.W., "Computational Construction Kits for Geometric Modeling and Design (Panel)," *ACM Symposium on Interactive 3D Graphics (I3D)*, March 2001, (TR2001-02).

Weiss, Y., Freeman, W.T., "On the Optimality of Solutions of the Max-Product Belief-Propagation Algorithm in Arbitrary Graphs," *IEEE Transactions on Information Theory*, February 2001.

Yedidia, J.S., "An Idiosyncratic Journey Beyond Mean Field Theory," *Advanced Mean Field Methods, Theory and Practice*, ISBN: 0-262-15045-9, pps 21-36, February 2001, (TR2000-27).

Marks, J.W.; Freeman, W.T.; Leitner, H., "Teaching Applied Computing without Programming: A Case-Based Introductory Course for General Education," *ACM Technical Symposium on Computer Science Education (SIGCSE)*, ISBN: 1-58113-329-4, Vol 33, Issue 1, pps 80-84, February 2001, (TR2001-06).

Divakaran, A.; Peker, K.A.; Sun, H., "Video Summarization Using Motion Descriptors," *SPIE Conference on Storage and Retrieval for Media Databases*, January 2001.

Peker, K.A.; Divakaran, A.; Papathomas, T.V., "Automatic Measurement of Intensity of Motion Activity of Video Segments," *SPIE Conference on Storage and Retrieval for Media Databases*, Vol 4315, p 341-351, January 2001.

Davies, J.R.; Gertner, A.; Lesh, N.B.; Rich, C.; Rickel, J.; Sidner, C.L., "Incorporating Tutorial Strategies into an Intelligent Assistant," *ACM International Conference on Intelligent User Interfaces*, ISBN: 1-58113-325-1, pps 53-56, January 2001, (TR2000-30).

2000

Freeman, W.T.; Weiss, Y., "On the Fixed Points of the Max-Product Algorithm," *IEEE Transactions on Information Theory*, 2000, (TR99-39).

Strohecker, C., "Cognitive Zoom: From Object to Path and Back Again," *Spatial Cognition II*, Vol 1849, pps 1, 2000, (TR2000-04).

Vetro, A.; Divakaran, A.; Sun, H.; Poon, T.C., "Transcoding System Based on MPEG-7 Meta-Data," *IEEE Pacific-Rim Conference on Multimedia (PCM)*, December 2000.

Fisher III, J.W.; Darrell, T.; Freeman, W.T.; Viola, P.A., "Learning Joint Statistical Models for Audio-Visual Fusion and Segregation," *Advances in Neural Information Processing Systems (NIPS)*, Vol 13, December 2000.

Moghaddam, B.; Yang, M-H., "Sex with Support Vector Machines," *Advances in Neural Information Processing Systems (NIPS)*, Volume 13, December 2000.

Yedidia, J.S.; Freeman, W.T.; Weiss, Y., "Generalized Belief Propagation," *Advances in Neural Information Processing Systems (NIPS)*, Vol 13, pps 689-695, December 2000, (TR2000-26).

Moghaddam, B.; Jebara, T.; Pentland, A., "Bayesian Face Recognition," *Pattern Recognition*, Vol 33, Issue 11, pps 1771-1782, November 2000, (TR2000-42).

Ogata, M.; Pfister, H.; Lauer, H.C.; Dohi, Y., "A Parallel Pipeline Convolution for Perspective Projection in Real-Time Volume Rendering," *The Journal of the Institute of Image Information and Television Engineers*, Vol 54, Issue 9, November 2000.

Shen, C.; Mizunuma, I., "RT-CRM: Real-Time Channel-Based Reflective Memory," *IEEE Transactions on Computers*, November 2000, (TR2000-41).

Chung, S-T; Gonzalez, O.; Ramamritham, K.; Shen, C., "CReMeS: A CORBA Compliant Reflective Memory Based Real-Time Communication Service," *IEEE Real-Time Systems Symposium (RTSS)*, pps 47-56, November 2000, (TR2000-18).

Bober, M.; Asai, K.; Divakaran, A., "MPEG-4/7 Based Internet Video and Still Image Browsing System," *SPIE Conference on Multimedia Systems and Applications*, Vol 4209, November 2000.

Horng, J.H.; Bao, J., "An Adaptive Diversity Combining Digital Receiver for Multiple Multiplexing Schemes," *IEEE International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, November 2000.

Gu, D.; Bao, J., "Performance Evaluation of a High-Speed Data Transmission Scheme with Modulo-Type Pre-equalization," *Asilomar Conference on Signals, Systems and Computers*, Vol 2, pps 1613-1616, October 2000.

Freeman, W.T.; Pasztor, E.C.; Carmichael, O.T., "Learning Low-Level Vision," *International Journal of Computer Vision*, Vol 40, Issue 1, pps 24-57, October 2000, (TR2000-05).

Strohecker, C.; Slaughter, A.H.; Horvarth, M.A.; Appleton, N.J., "Zyklodeon: A Software Construction Kit Modeling Cyclic Timing Patterns," *ACM International Multimedia Conference (MM)*, ISBN: 1-58113-311-1, pps 31-34, October 2000, (TR2000-29).

Darrell, T.; Fisher III, J.W.; Viola, P.A.; Freeman, W.T., "Audio-Visual Segmentation and 'The Cocktail Party Effect'," *International Conference Advances in Multimodal Interfaces (ICMI)*, Vol 1948, pps 32-40, October 2000.

Bhalerao, A.; Pfister, H.; Halle, M.; Kikinis, R., "Fast Re-Rendering of Volume and Surface Graphics by Depth, Color and Opacity Buffering," *Journal of Medical Image Analysis*, Vol 4, Issue 3, pps 235-251, September 2000.

Lesh, N.B.; Marks, J.W.; Patrignani, M., "Interactive Partitioning," *International Symposium Graph Drawing*, Vol 1984, pps 31-36, September 2000, (TR2000-31).

Divakaran, A., Peker, K.A.; Sun, H., "A Region Based Descriptor for Spatial Distribution of Motion Activity for Compressed Video," *IEEE International Conference on Image Processing (ICIP)*, Vol 2, pps 287-290, September 2000.

Divakaran, A., Vetro, A., Asai, K.; Nishikawa, H., "Video Browsing System Based on Compressed Domain Feature Extraction," *IEEE Transactions on Consumer Electronics*, August 2000.

Brand, M.E.; Kettnake, V., "Discovery and Segmentation of Activities in Video," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, August 2000.

Casey, M.A.; Westner, W., "Separation of Mixed Audio Sources by Independent Subspace Analysis," *International Computer Music Conference (ICMC)*, August 2000, (TR2001-31).

Lin, I.; Divakaran, A.; Vetro, A.; Kung, S.Y., "Efficient Representation and Comparison of Multimedia Content Using DAG-Composition," *IEEE International Conference on Multimedia and Expo (ICME)*, Vol 2, pps 895-898, July 2000.

Anderson, D., Anderson, E., Lesh, N.B., Marks, J.W, Mirtich, B., Ratajczack, D.; Ryall, K., "Human-Guided Simple Search," *National Conference on Artificial Intelligence (AAAI)*, ISBN 0-262-5112-6, August 2000, (TR2000-16).

Brand, M.E.; Hertzmann, A., "Style Machines," *ACM SIGGRAPH*, ISBN: 1-58113-208-5, pps 183-192, July 2000, (TR2000-14).

Frisken, S.F.; Perry, R.N.; Rockwood, A.P.; Jones, T.R., "Adaptively Sampled Distance Fields: A General Representation of Shape for Computer Graphics," *ACM SIGGRAPH*, ISBN: 1-58113-208-5, pps 249-254, July 2000, (TR2000-15).

Anderson, D.; Frankel, J.L.; Marks, J.W.; Agarwala, A.; Beardsley, P.A.; Hodgins, J.K.; Leigh, D.L.; Ryall, K.; Sullivan, E.; Yedidia, J.S., "Tangible Interactions and Graphical Interpretation: A New Approach to 3D Modeling," *ACM SIGGRAPH*, ISBN: 1-58113-208-5, pps 393-402, July 2000, (TR2000-13).

Mirtich, B., "Timewarp Rigid Body Simulation," *ACM SIGGRAPH*, ISBN: 1-58113-208-5, pps 193-200, July 2000, (TR2000-17).

Pfister, H.; Zwicker, M.; van Baar, J.; Gross, M., "Surfels: Surface Elements as Rendering Primitives," *ACM SIGGRAPH*, ISBN: 1-58113-208-5, pps 335-342, July 2000, (TR2000-10).

Matusik, W.; Buehler, C.; Raskar, R.; Gortler, S.J.; McMillan, L., "Image-Based Visual Hulls," *ACM SIGGRAPH*, ISBN: 1-58113-208-5, pps 369-374, July 2000.

Project Reports

The body and soul of any research lab is the portfolio of projects it pursues. Therefore it is appropriate that the main body of this annual report consists of descriptions of the various projects being done at MERL. For ease of reference, the reports are grouped into ten topic areas.

- Advanced Digital Television
- Artificial Intelligence
- Audio Video Processing
- Collaborative and Interaction
- Computer Vision
- Digital Communications
- Graphics
- Net Services
- Networks
- Spoken Language Interfaces

Each topical section begins with a short discussion of the topic area, highlighting MERL's major efforts. It then continues with a number of one-page project reports. These reports describe projects completed in the last twelve months and major milestones in continuing efforts. The individual project reports begin with a brief summary at the top, followed by a more detailed discussion. The bottom of the report indicates the principal lab at MERL involved with the project and a contact person. Also included is a characterization of the type of project. The purpose of this is to indicate the kind of result that has been obtained.

- Initial Investigation – Work is underway on the project, but no firm results have been obtained yet. The project report is included to give a better understand of a direction in which MERL is heading.
- Research – The results obtained are in the form of papers, patents, and/or research prototypes. They represent valuable knowledge, but significant advanced development work will be required before this knowledge can be applied to products.
- Advanced Development – The results are (or will be) in forms that can be directly used in product development. The exact form of the result depends on what is being produced. For software projects, the results are typically code that can be directly used in products. For semiconductor chip projects, the results are typically in the form of detailed specifications for algorithms to be embedded in silicon.

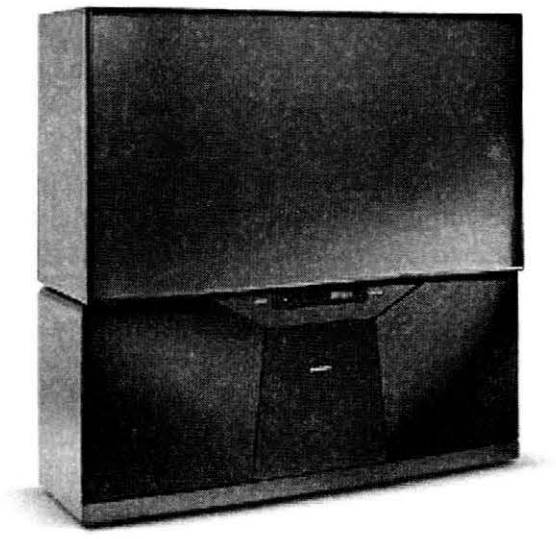
Advanced Digital Television

Advanced Digital Television opens our eyes to a whole new dimension in home entertainment. Beyond the breath taking picture quality rivaling film theaters, it is destined to become the information center of our home. Seamlessly integrating all the digital information in the house and becoming the heart of a network of digital home appliances.

At MERL, we have been carefully nurturing this budding new technology since its inception, providing key enabling know-how to help Mitsubishi Electric provide the ultimate digital television for our customers. Realizing that as digital television gradually come into our living rooms, we will increasingly turn to it not just for entertainment, but the management of all the digital appliances, too. With this vision, we now focus on lobbying the broadcast infrastructure for delivering more digital services as well as extending the networking and information management capability of the digital television itself, providing more digital content to satisfy the senses and managing the myriad network of digital appliances appearing in our homes.

Project Descriptions

2nd Generation Digital Television Receiver Module Development



Mitsubishi Digital Electronics America (MDEA) is the foremost leader in consumer digital televisions with the highest market share. To maintain this position, each generation of the new products must incorporate cutting-edge new technology combined with state of the art performance.

After introducing the industry's first digital television receiver with 400 megabits per second IEEE-1394 digital connection coupled with the powerful HAVi networking protocol, the 2003 Mitsubishi DTV will extend this successful platform with more features and better performance.

Background and Objectives: DTV performance is under constant improvement and the aspect that receives the most scrutiny is the terrestrial digital broadcast signal reception capability. This part of the DTV system for receiving the digital signal is called the "front end," consisting of a tuner and a demodulator. MERL has invented a new technique for improving the reception performance of the DTV and a patent is under preparation. MERL will further contribute to the design, evaluation, and field-testing of the front end culminating to a successful mass production of the second generation DTV receiver module. In the mean time, MERL also serves as the east coast base for investigating DTV reception problems reported by authorized dealers, most notably in the Philadelphia area and also Boise, Idaho. In these two cases, MERL was able to identify broadcast standards violations by the stations as the cause and resolved the problem without any modifications to our television sets.

Technical Discussion: The tuner and the demodulator each supplied by a different manufacturer need to be integrated and optimized to form a complete front end. The accompanying circuitry for this integration must be designed by each television manufacturer and may include new techniques that differentiate its DTV sets from that of the competitor's. A new signal processing technique proposed by MERL will improve the weak signal reception performance of the DTV. This technique can also be applied to almost any manufacturer's design and holds possible licensing opportunities.

Collaboration: Mitsubishi Digital Electronics America (MDEA)

Future Direction: MERL-MHL will continue to contribute to MDEA's effort of developing the second generation DTV receiver module with production starting in early 2003.

Contact: George Fang
http://www.merl.com/projects/dm2/

Lab: MERL Murray Hill
Project Type: Advanced Development

U.S. DTV on Cable

Since FCC's adoption of DTV standards in the U.S. in 1996, the number of DTV stations on the air has increased to 200+ stations by end of 2001, and has almost doubled that in the first half of 2002 alone. The numbers represent over 90% of U.S. households have access to at least one digital broadcast over the air. The cable companies, who control access to 70% of U.S. homes, are refusing to carry the DTV signals from broadcasters, citing a variety of technical problems. This study looks into the status of cable industry's analog-to-digital transition, in order to help MDEA's commitment to DTV

Background and Objectives: MDEA is the leading manufacturer of DTV sets in North America. Cable DTV standard is different compared to the broadcast DTV standard. The goal is to investigate how MDEA can take advantage of the digital potential from cable.

Technical Discussion: Cable allows system information to be delivered out-of-band, which is a dramatic departure from the terrestrial broadcast. An out-of-band transmission means DTV receivers must also monitor the signal from another frequency, and thus increases the complexity of the receiver design. Document SCTE/DVS 234 is an approved document to standardize the scheme for out-of-band system information deliveries. SCTE/DVS 241 on the other hand describes the transport layer characteristics and normative specifications of the in-band service multiplex and transport system standard for cable DTV. One of the more important standard on the cable is SCTE/DVS 295, which defines the characteristics and normative specifications for the interface between Point of Deployment (POD) security modules owned and distributed by cable operators, and commercially available consumer receivers and set-top terminals (host devices) that are used to access multi-channel television programming carried on North American cable systems. In addition SCTE/DVS 301 further defines the POD standard to include copy protection for content delivery to the receiver, with a goal of preventing unrestricted copying of high valued contents. Most recently SCTE/DVS 311 was developed to describe the methods to transmit multicast IP data over MPEG 2 digital transport streams.

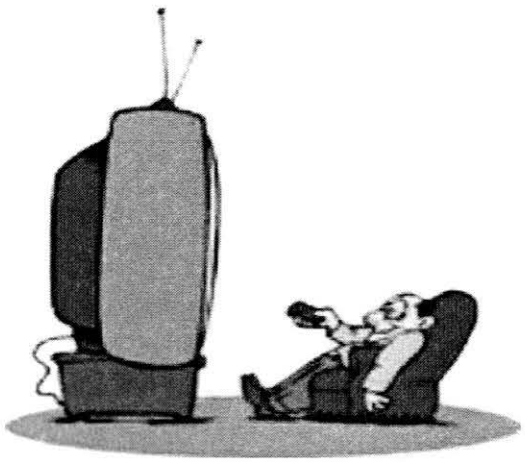
Collaboration: Society of Cable Telecommunications Engineers/Digital Video Subcommittee, Mitsubishi Digital Electronics America, Inc., Advanced Television Systems Committee.

Future Direction: Continue investigation in the digital cable transition will help MDEA's commitment in the DTV transition.

Contact: James Fang
http://www.merl.com/projects/us_dtv_on_cable/

Lab: MERL Murray Hill
Project Type: Advanced Development

U.S. DTV Broadcasting Implementation Issues



The progress of analog to digital television transition is coming along at an exponential rate. As more and more stations start to broadcast DTV signals, it is apparent that certain problems are starting to catch the broadcast industry's attention. The implementation issues are being identified and possible solutions are proposed with experts from the broadcast industry and equipment manufacturers that include both broadcast and consumer devices. Participation in the standardization process and investigation of problems/solutions will greatly benefits advanced product development.

Background and Objectives: As broadcast standards are approved, many implementation issues arise when actually trying making the system work. Some of those approved standards then go through a period of evolution before the actual deployment.

Technical Discussion: FCC has a deadline for consumer DTV manufacturers to include DTVCC (digital television closed caption) decoding circuitry, but there isn't any test signals for the manufacturers to test the decoder. Meanwhile efforts in developing broadcast DTVCC encoders are slow in coming, given FCC also has a deadline for broadcasters to include DTVCC data in DTV broadcast. Ambiguities in the DTVCC standard also took some time to clear. Progress is slow in this area of work. Latency and timing: Digital signal has a different path than analog, and the end result is program starts and ends at different times compared to the announced times. This represents great inconveniences to the viewers particularly if one tries to record successive programs from different stations. From the broadcast standpoint, signal delay includes distribution, plant routing, conversion, switching, transmission, and encoding. DCC (directed channel change) has been approved as a standard for over 2 years. Implementation issues include seamless transition between channel changes that include both audio and video, and the triggering mechanism that initiates the actual channel change. Another area of study is the enhancements to the 8VSB modulation standard as adopted by the FCC. The possibility of changing the modulation scheme inspired investigation of the transport standard. A different transport standard must be developed before FCC's decision. Other issues include advanced EPG (electronic program guide), RF link budget, and standard studio equipment interfaces.

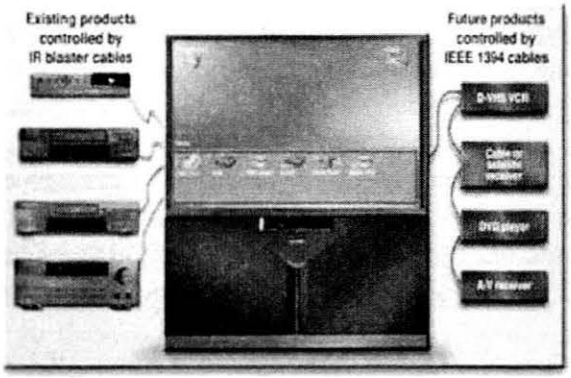
Collaboration: Advanced Television Systems Committee; Consumer Electronics Association; Mitsubishi Digital Electronics America; National Center for Accessible Media.

Future Direction: Any help in identifying implementation issues will assist MDEA in developing new products for the market.

Contact: James Fang
http://www.merl.com/projects/us_dtv_broadcasting/

Lab: MERL Murray Hill
Project Type: Advanced Development

HAVi and IEEE 1394 Standard Activities



In the near future, Consumer Electronics devices will be interconnected to facilitate shared access to content, services, and control & command functions. The HAVi goal is to facilitate plug & play interoperability of AV devices using the IEEE 1394 cable interface. Typical home devices targeted by HAVi include DTV, VCR, DVD, and Camcorders. HAVi addresses the complete set of requirements for networked AV equipment including: discovery of connected devices, description of capabilities, control, and most importantly, delivery of content (usually MPEG content) with guaranteed QoS implicitly done by adopting IEEE 1394.

Background and Objectives: MDEA introduced the first HAVi products (HDTV and HD D-VHS VCR) to the market in 2001. The main objective is to conduct research on innovative networking areas for introduction into the various standardization processes such as HAVi, IEEE 1394.x and 1394TA.

Technical Discussion: IEEE 1394 provides a good networking foundation for Consumer Electronics devices. Hot plug and play, high data rate, and QoS support are among others, the key benefits. Being IEEE1394 a specification for layers 1,2 and 3, it does not provide any interoperability guarantee among devices implementing dissimilar higher layers. As an attempt to solve this problem, HAVi is a framework built on top of IEEE 1394 and complemented by other AV specifications like IEC 61883 and AV/C. HAVi enables AV devices to interoperate by defining a set of generic APIs which cover a great percentage of current devices. Future proof is achieved by the support of byte-code (HAVi adopted Java) that can run on any device. Future devices will simply contain their own byte-code that can be uploaded into other devices in the network without the need for any device to have any a priori knowledge. Once the connectivity problem among AV devices is solved, the next step is connectivity to other clusters and/or external networks.

Collaboration: MDEA (Mitsubishi Digital Electronics America)

Future Direction: CE devices will gradually depend on connectivity inside and outside the home to gather a wide variety of data ranging from content and metadata (e.g. EPG, CD song titles, etc.) to highly personalized information such as time/location/environment-aware data and alerts (see related PAMLink projects). In this respect, new research topics include IPv4/IPv6 over 1394, Bridging (e.g. 1394 to 1394, 1394 to Internet), and a File Management system for HAVi.

Contact: Fernando Matsubara
http://www.merl.com/projects/havi1394/

Lab: MERL Murray Hill
Project Type: Advanced Development

Artificial Intelligence

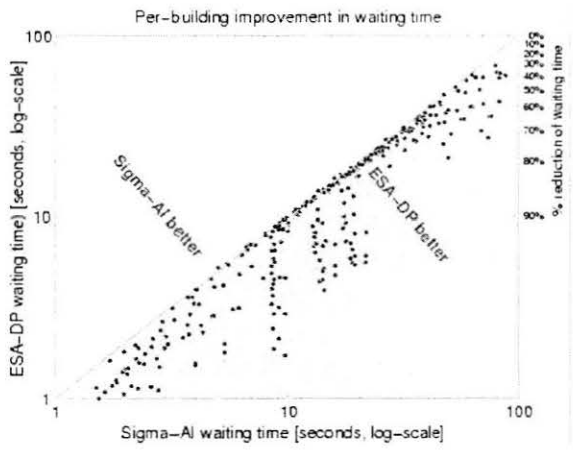
"Artificial intelligence" deals with intractable problems, where the amount of work a computer must do grows exponentially in the size of the problem. Historically, the name derives from the fact that many problems associated with human intelligence have this property, including language understanding, perception, making decisions from uncertain evidence, and planning. In all cases, there is exponential growth in the number of possible solutions that must be considered. In recent years, AI's mathematical foundations and engineering methods have rapidly matured. In particular, a new emphasis on high-quality approximate solutions that can be efficiently computed has led to practical algorithms for economically significant problems that were previously difficult for both people and machines.

MERL has a growing international reputation for its theoretical and practical results in machine learning, computer vision, Bayesian inference algorithms, coding and compression, and planning. MERL AI projects typically focus on the interface between man and machine – vision, speech interpretation, graphics, and user interfaces – where the ability to learn and improve with experience is key to dealing with issues of intractability.

Recent MERL projects in vision, speech, and graphics are each covered in separate sections. Pages 53 & 54 describe a software infrastructure and methodology for supporting human-computer collaboration where the computer acts as an intelligent partner. Several year 2001 AI projects yielded substantial advances in classic industrial problems: Optimal control of groups of elevators in tall buildings (page 50); data-mining extremely large data-sets to predict consumer behavior (page 51); and a new class of algorithms for overconstrained inference problems such correcting errors in data transmitted over a noisy channel (page 55).

Project Descriptions

Optimal Group Elevator Control



In elevator scheduling, the goal is to assign elevators to pick up passengers while minimizing the waiting times of all passengers in the building. We have developed a unified approach for making finite-horizon-optimal decisions in all types of traffic (morning up-peak, midday mixed, evening down-peak). In benchmark tests against MELCO's state-of-the-art Sigma-AI 2200 elevator controller, our new ESA-DP (Empty the System Algorithm, Dynamic Programming) controller reduces waiting times by 10%-80% in heavy and light traffic, and by 5%-15% in medium traffic. Harder scheduling problems yield larger gains. The ESA-DP controller substantially increases the thoughput of an elevator system. Equally important, the ESA-DP controller can profitably exploit new sources of information, for example, passenger counting via camera or early indications of passenger destinations from extended call panels.

Background and Objectives: One of the main selling points of an elevator system is how quickly it responds to passenger requests. However, the problem of estimating response time was thought to be intractable, so historically elevator control algorithms have been designed to optimize various approximations to response time. We set out to develop an efficient method to explicitly compute and optimize response time.

Technical Discussion: To compute response time, one must consider all possible elevator itineraries in the future. This was thought to be intractable because passenger destinations are mostly unknown, and therefore an exponential number of possibilities must be considered. We developed a way to fold all possible itineraries into a dynamic programming trellis and evaluate the trellis in linear time, thereby computing the expected response time for all passengers over all possible scenarios. On this basis, we select a car-to-passenger assignment that is optimal up the time where the system would empty itself of all current waiting passengers. Similar methods allow us to position idle cars in a way that optimally anticipates future passenger arrivals. Despite that fact the computation must consider millions or even billions of possible itineraries, each decision takes roughly 5 milliseconds.

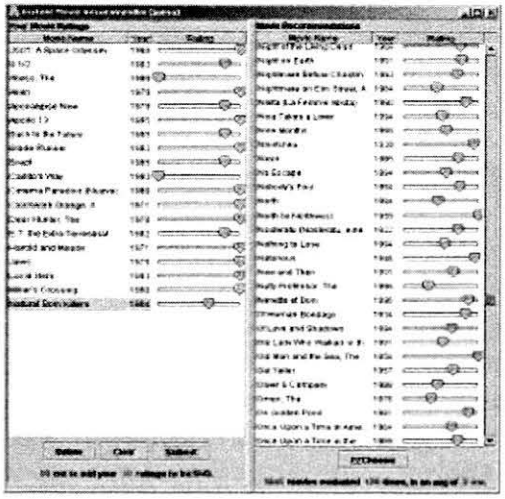
Collaboration: This work is being done with the help and advice of Kouichi Sasakawa and the building systems group of the Industry Solutions Technology department, Sanken.

Future Direction: Minimizing the waits of passengers who have not yet called for an elevator.

Contact: Matthew Brand
http://www.merl.com/projects/ElevatorControl/

Lab: MERL Cambridge Research
Project Type: Research

Data-mining and Recommending



The Instant Movie Recommender (IMR), pictured left, predicts how you would rate over 1000 movies on the basis of your ratings of a few familiar ones. As you vary the rating of one movie, all the other movies are instantly re-ranked and re-sorted. Predictions are made on the basis of correlations between ratings by previous users—people who like "X" tend not to like "Y", etc. The IMR "learns" from your ratings—updating the correlational model "on the fly". This update is a form of data-mining, but with substantially better economics and usability than the standard practice of warehousing and batch-processing data.

The IMR showcases a new technology for fast updating of a Singular Value Decomposition— a decomposition of tabular data into simple factors. The technology is distinguished both by its speed—it is the first linear-time algorithm—and by its ability to handle tables with many missing elements—a common problem in data-mining.

Background and Objectives: The SVD forms the core of most data-mining algorithms and thousands of algorithms in signal processing. The "thin" SVD decomposes tabular data into the product of two small matrices, and is very useful for data compression, noise suppression, and prediction. For very large tables, computing an SVD is impractical because the compute time grows quadratically with the size of the table. Worse, the SVD is not uniquely defined if the table is missing some entries. We set out to develop an SVD algorithm that is suitable for data-sets that are far too large to fit into the computer's memory and that are missing many elements.

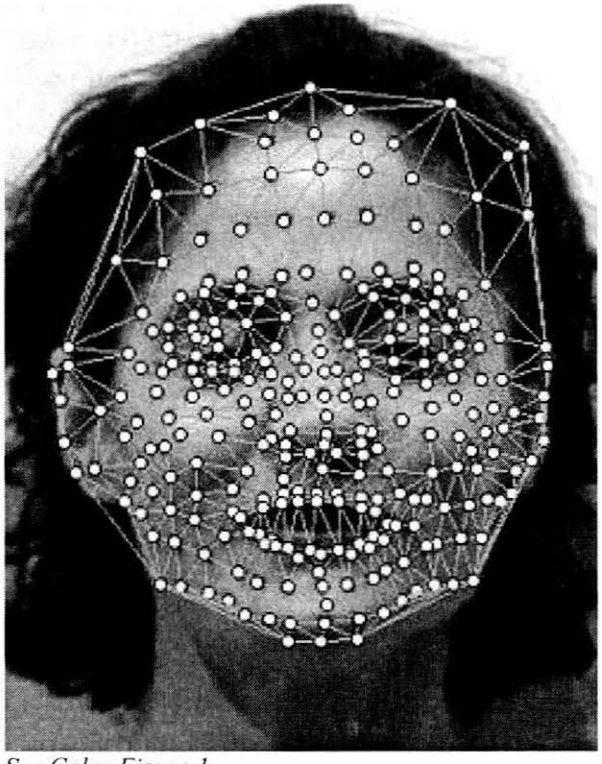
Technical Discussion: The new Incremental Imputative SVD allows an SVD to be computed from streaming data. The data table arrives one row or column at a time, and the SVD is updated to reflect the newly arrived information. The data need not be stored. If the row is missing entries, the algorithm chooses values that are most likely to result in good prediction of future row. This is how the IMR predicts ratings of unseen movies. Experiments with a dataset used by the data-mining community for benchmarking indicate that the IMR is quite accurate, making predictions within 2 rating points of the "true rating" more than 99% of the time.

Collaboration: The IMR is being developed into a candidate MELCO product with Shinsuke Azume (Sanken).

Future Direction: We are now studying the problem of efficiently downdating the SVD (retracting ratings).

Contact: Matthew Brand, Frederick J. Igo, Jr., David Wong **Lab:** MERL Cambridge Research
http://www.merl.com/projects/DataMining/ **Project Type:** Initial Investigation

Incremental SVD of Incomplete and Uncertain Data



See Color Figure 1

The incremental incomplete singular value decomposition makes it possible to produce useful inferences from data even when most the data is missing. For example, we have used it to accurately predict consumer preferences (what movies one would like) from a table of movie ratings that is 97% empty.

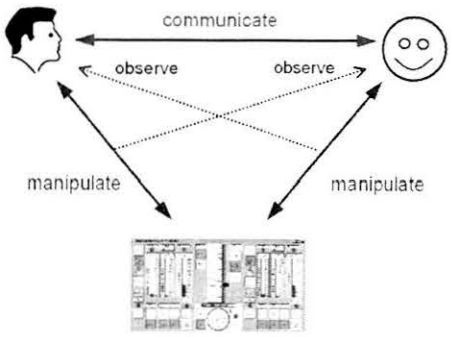
Background and Objectives: The singular value decomposition (SVD) forms the core of thousands of algorithms in signal processing. Computing an SVD becomes difficult or even impossible if some of the data is missing, if the dataset is too large, or if the data is contaminated with non-white noise. We set out to develop algorithm for the SVD that can be computed online as the data comes in and can handle varying noise conditions, including datapoints that are entirely missing.

Technical Discussion: The incremental incomplete SVD combines and improves recently developed methods for updating an SVD with new data, imputing missing values, and finding most probable least-squares solutions when noise is anisotropic and inhomogeneous. Remarkably, the resulting algorithm has better time and space complexity than batch SVD algorithms and computes solutions of comparable accuracy. To date, the algorithm has been used to predict consumer preferences, to extract 3D shape information from video, and compute compression bases for audio and video datasets having tens of millions of datapoints.

Contact: Matthew Brand
http://www.merl.com/projects/IISVD/

Lab: MERL Cambridge Research
Project Type: Research

COLLAGEN: Java Middleware for Collaborative Interface Agents



COLLAGEN (for COLLaborative AGENt) is Java middleware for building collaborative agents. A collaborative agent is a software program that helps users solve problems, especially in complex or unfamiliar domains, by correcting errors, suggesting what to do next, and taking care of low-level details. A collaborative agent can be added to an existing graphical user interface, such as a software simulator, or integrated into the design of a new hardware device, such as a personal video recorder (PVR). COLLAGEN is currently being used to build prototype systems for a range of applications, including power plant operator training, car navigation, and spoken-language web form filling.

Background and Objectives: The theoretical foundations of COLLAGEN derive from the study of naturally occurring human collaboration, such as two people assembling a complex mechanical device or two computer users working on a spreadsheet together. The practical objective of the project is to maximize the software reuse in building collaborative agents for many different applications.

Technical Discussion: As shown in the accompanying diagram, a collaborative agent, in general, communicates with the user (using either natural or artificial language), manipulates some shared hardware or software artifact, and can observe the user's manipulations of that artifact. The heart of COLLAGEN is its representation of the current state of a collaborative dialogue, consisting of a plan tree, which tracks the status of the shared domain tasks, and a focus stack, which tracks the current focus of attention. COLLAGEN automatically updates these data structures whenever either the user or the agent speaks or performs a manipulation. The agent then refers to these data structures to determine what to do or say in response. To apply COLLAGEN in particular domain, such as power plant operation or PVR programming, the key information a developer must provide is the task model, which is an abstract formalization of the hierarchical structure of the typical tasks and subtasks in the domain.

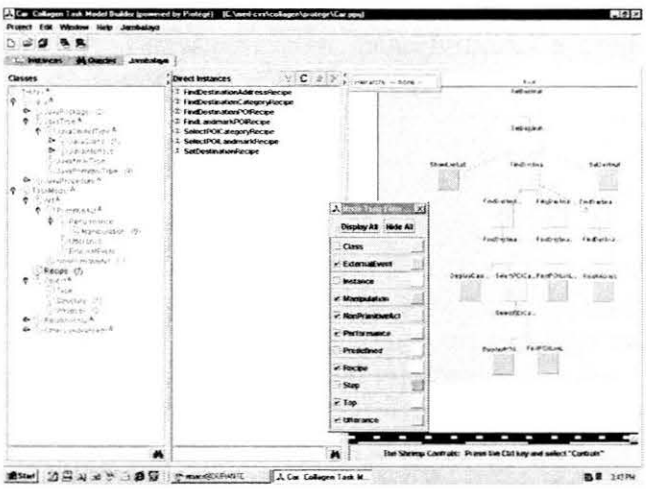
Collaboration: In collaboration with Mitsubishi Electric's Advanced Technology R&D Center, Information Technology R&D Center, Automotive Electronics Development Center, and Energy and Industrial Systems Center, we have developed, or are developing, prototype collaborative agents using COLLAGEN for power plant operator training, car navigation, and spoken-language web form filling.

Future Direction: In addition to continuing to seek new applications (especially involving speech), we plan improvements in the basic operation of COLLAGEN in the areas of turn taking, causal knowledge, and negotiation.

Contact: Charles Rich
http://www.merl.com/projects/collagen/

Lab: MERL Cambridge Research
Project Type: Research

A Tool for Building and Visualizing Task Models



A task model is an abstract formalization of the hierarchical structure of the typical tasks and subtasks in a domain. Building a task model is an essential step in the development of many kinds of intelligent systems, including collaborative agents. Building a task model can also be quite labor-intensive without adequate tool support. We have implemented a Java-based tool, called COLLTASK, for building and visualizing task models, which can be used together with the COLLAGEN middleware to facilitate the implementation of collaborative agent systems.

Background and Objectives: This work is part of the process of transitioning COLLAGEN from a research prototype to a practical technology for building collaborative agents. COLLTASK replaces or supplements COLLAGEN's current textual representation for task models with a modern, graphically-oriented tool.

Technical Discussion: COLLTASK is implemented in Java using Protege-2000 (http://protege.stanford.edu), which is an open-source platform for knowledge-acquisition tools. For the graphical visualization capabilities, we have also incorporated Jambalaya (http://www.csr.uvic.ca/shrimpviews), which is free software for research use. COLLTASK can both read and write COLLAGEN's textual representation for task models. It can also read and write task models in several generic formats, such as RDF and JDBC.

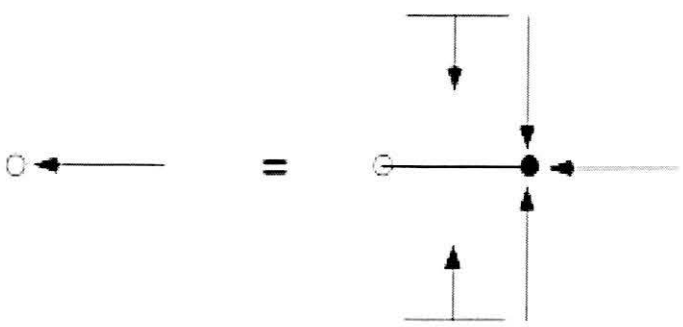
Collaboration: Mitsubishi Electric's Advanced Technology R&D Center supported this work.

Future Direction: The current COLLTASK tool is a first prototype. We plan to continue to improve it as we gain experience with using it. In addition, we expect to benefit when the developers of Jambalaya complete a closer integration of Jambalaya with Protege-2000, which is currently under way. A more fundamental improvement to COLLTASK will be to incorporate techniques we have developed for automatically learning task models from example traces of domain actions (learning by demonstration).

Contact: Charles Rich, Andrew Garland
http://www.merl.com/projects/colltask/

Lab: MERL Cambridge Research
Project Type: Advanced Development

Generalized Belief Propagation Algorithms



Many problems in computer vision, machine learning and inference, diagnosis, statistical physics, error-correcting coding, and combinatorial optimization can be posed in terms of a probabilistic graphical model consisting of a lattice of nodes with links connecting nodes that influence each other. Typically, one asks for the probability that a given node or collection of nodes is in some state, given the states of another set of nodes. We have developed a new class of algorithms that can often solve such problems much more quickly and accurately than previously-known algorithms.

Background and Objectives: Our generalized belief propagation (GBP) algorithms generalize and improve previously developed "belief propagation" (BP) algorithms, and give more accurate results with fewer convergence problems. Applied to the specific application of decoding error-correcting codes, we have shown that GBP decoding outperforms standard BP decoding for codes whose graphical representation contains short cycles. Since standard BP decoding has proven to be remarkably powerful for such codes as turbocodes and low-density parity check codes, this is a potentially exciting result.

Technical Discussion: GBP algorithms are theoretically based on the insight that their fixed points are equivalent to the stationary points of an approximate region-based "free energy." We have developed a general theory for constructing good free energy approximations, based on a "region graph" approach. In this approach, the region graph also indicates directly which regions of nodes send information to other regions of nodes in the corresponding GBP algorithm. The standard BP algorithm and the GBP algorithms that we previously developed based on the cluster variational method (giving Kikuchi approximations to the free energy) are special cases of those developed with the more general region graph approach. An important advantage of the region graph approach is that it gives guidance about how to construct GBP algorithms that can be expected to give good results without excessive computational cost.

Collaboration: with Yair Weiss (Hebrew University, Jerusalem), Bill Freeman (MERL and MIT), Marc Fossorier (University of Hawaii), Jinghu Chen (MERL and University of Hawaii)

Future Direction: We are now implementing GBP decoding algorithms for powerful error-correcting codes that cannot be decoded using the standard BP decoding algorithm. These algorithms, developed using the region graph approach, should be sufficiently efficient to be practical for real-world applications.

Contact: Jonathan Yedidia, William Freeman
http://www.merl.com/projects/GBPalgorithms/

Lab: MERL Cambridge Research
Project Type: Research

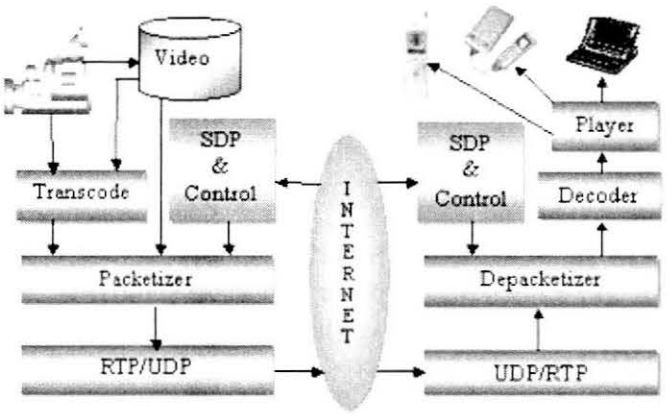
Audio Video Processing

Audio-Video Processing in general addresses reproduction of the audio-video content on one hand and analysis of the content on the other. In other words, audio-video processing is concerned with producing better-looking pictures and crisper audio as well as with understanding the audio-visual content. Automatic Speech Recognition is a prominent example of audio-visual processing applied to content understanding. While it pervades our daily life, it continues to a vibrant field rich with research challenges.

The Audio Video Processing effort at MERL consists of MPEG-4, MPEG-7, and related activities. We have successfully proposed our technologies for adoption into these standards. Our object-based rate control algorithm is now part of the informative (non-normative) part of the MPEG-4 standard. Our Video Motion Activity Descriptor, Directed Acyclic Graph based Description Scheme, Video Transcoding Description Scheme, and Audio Content Indexing and Extraction Method are all now part of the normative part of the MPEG-7 draft international standard. We are now pursuing various applications of our existing technology as well as developing new technology. Our efforts consist of both developing individual technologies as well as innovative combinations of technologies. Examples of projects that combine our different technologies include the Video Transmission Platform, that combines MPEG-2 and MPEG-4 technology and the MPEG-7 Video Browsing and Summarization that combines audio and video indexing and summarization. Examples of projects that capitalize on individual video technologies include the MPEG Transcoding for Surveillance which consists of video conversion from MPEG-2 to MPEG-4, MPEG Transcoding for DVD systems which consists of video conversion from high bitrate MPEG-2 to lower bitrate MPEG-2, and Video Object Segmentation which consists of automatic segmentation of video into its constituent objects. Examples of projects that capitalize on audio techniques include MPEG-7 Sound Recognition and MPEG-7 Music Player, which helps find music using MPEG-7 technology. Our projects thus cover both key aspects of audio-video processing i.e. Content Reproduction and content understanding.

Project Descriptions

Video Transmission Platform



People have desired end-to-end visual communications for many years. However, real video communications over the Internet has not been widely and successfully used in our daily life because the current best-effort Internet does not offer any quality of service (QoS) guarantees. We have developed a client/server system that allows video transmission over heterogeneous networks. We aim to use this platform for further research on QoS solutions.

Background and Objectives: There are a few companies, such as, Apple, Microsoft and RealNetworks providing platforms for video streaming. Unfortunately, those platforms use different transport schemes and media types, incompatible to each other. The recently released Internet Streaming Media Alliance Implementation Specification (ISMA) offers an open standard for media streaming with the goal to use the existing open standards that people can follow to build interoperable audio and video systems for use on IP networks and the Internet. Our current platform is fully compliant with the ISMA specification. Our ongoing work is to enhance the functionalities of the platform to allow users to interactively access and control the video transmission in real-time.

Technical Discussion: This platform takes MPEG-4 elementary streams (ES), MP4 file and MPEG-2 transport streams (TS) as its input. In the case of MPEG-2 TS, we have integrated a transcoder that converts the MPEG-2 TS to MPEG-4 ES. Clients and server initiate their session connections through a session description protocol (SDP). RTP (real-time protocol), which provides the basic functionalities, such as sequence number, time stamps and multicasting is used for carrying real-time data over packet networks. The packetization is based on RFC 3016 RTP payload format for MPEG-4 video streams. At the client side, we are able to playback received MPEG-4 bitstreams on both Win32 and WinCE platforms.

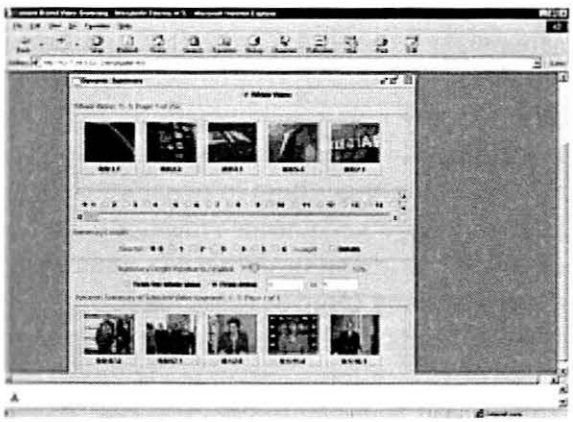
Collaboration: This project is performed in collaboration with the Multimedia Information Coding & Transmission Technology Department at Johosoken.

Future Direction: An RTSP (real-time streaming protocol) mechanism for negotiating QoS for video streams and an RTP/RTCP retransmission with forward error correction (FEC) mechanism will soon be implemented into the platform.

Contact: Hao-Song Kong, Anthony Vetro
http://www.merl.com/projects/video_transmission/

Lab: MERL Murray Hill
Project Type: Advanced Development

MPEG-7 Video Browsing and Summarization



The Murray Hill Laboratory (MHL) of Mitsubishi Electric Research Laboratories has developed a Motion-based Indexing and Summarization system for video. It is based on our work on descriptors of motion activity and their combination with descriptors of other features such as color, as well as our work on Directed Acyclic Graph (DAG) based Description schemes, both of which have also been accepted into the MPEG-7 standard. Most recently we have also incorporated audio features into our indexing framework. The purpose of the system is to help the user quickly traverse recorded video content using both top-down i.e. summarization based, and bottom-up, i.e. indexing based, access. We illustrate the MISE web-based interface in the figure on the left.

Background and Objectives: As more and more audio-visual content becomes available in digital form in various places around the world, the ability to locate desired content will become more and more important. Already text based search engines help retrieve textual data from the World Wide Web, but no equivalent identifying information exists for A/V content. The proposed MPEG-7 standard will standardize a multimedia content description interface that will enable efficient searching and browsing of worldwide multimedia content. In this project we emphasize the Personal Video Recorder application, that provides the user with the content he wants when he wants it by storing a large volume of content recorded from broadcast and then providing effective navigation of the stored content using summarization and indexing.

Technical Discussion: The system relies on extraction of compact descriptors in the compressed domain, which makes both the content preparation and the content access fast. It primarily relies on the MPEG-7 motion activity descriptor, and also makes use of simple color histograms. We have a unique motion activity based approach to video summarization.

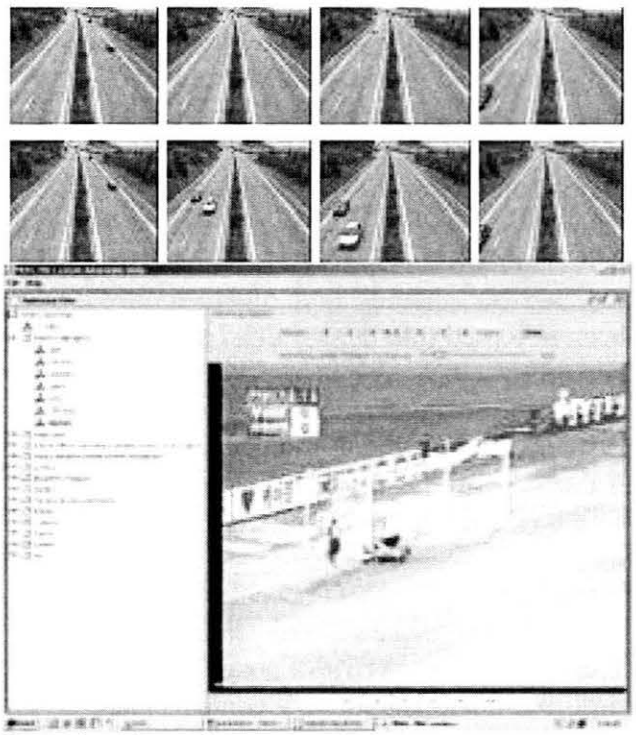
Collaboration: Joho-Soken, Sentan-Soken

Future Direction: Our current research focus is on Content Summarization. We are still working on improving and enhancing the summarization. Our target applications include Personal Video Recorders, Consumer Video Browsing for DVD players and other CE devices, Remote access of video, Surveillance etc.

Contact: Ajay Divakaran
http://www.merl.com/projects/video-browsing/

Lab: MERL Murray Hill
Project Type: Advanced Development

Event Detection



See Color Figure 2

The Murray Hill Laboratory (MHL) of Mitsubishi Electric Research Laboratories has developed a suite of Video Indexing, Summarization and Segmentation techniques. We have obtained promising results by applying them to event detection in traffic video and soccer video. We have obtained promising initial results and now plan to develop more accurate and sophisticated techniques for event detection in traffic and indoors surveillance video. The evident application is in the area of surveillance such as detecting accidents through analysis of traffic surveillance video illustrated at left. We have promising preliminary results on detection of onsets of traffic jams using our MPEG-7 motion activity descriptor. In another project, we are jointly investigating extraction of semantic features from low-level features of soccer games with Columbia University.

Background and Objectives: The proposed MPEG-7 standard will standardize a multimedia content description interface that will enable efficient searching and browsing of worldwide multimedia content. In this project we will work on both analysis of stored video using searching and browsing techniques for event detection as well as on direct event detection using MPEG-7 based techniques for content understanding in combination with our video segmentation technology.

Technical Discussion: Our system is currently based on dynamic feature extraction in the compressed domain. It primarily relies on the MPEG-7 motion activity descriptor, and also makes use of simple color histograms. We have a unique motion activity based approach to video summarization. We plan to refine our techniques as well as customize them to traffic video analysis.

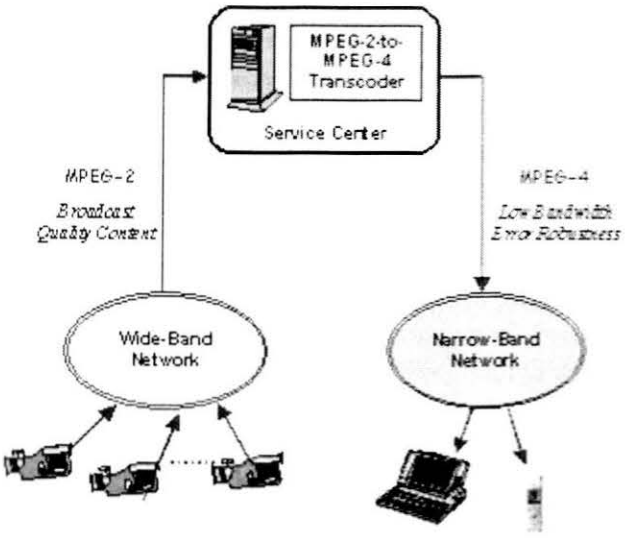
Collaboration: Sentan-Soken

Future Direction: Our current research focus is on deriving higher- level features from low level features using a combination machine learning techniques and domain knowledge. Our target applications are surveillance and video summarization for consumer video browsing.

Contact: Ajay Divakaran
http://www.merl.com/projects/event-detection/

Lab: MERL Murray Hill
Project Type: Advanced Development

MPEG Transcoding for Surveillance



In general, the purpose of a transcoder is to convert compressed content, such as an MPEG (Moving Pictures Experts Group) bitstream, into a format that satisfies transport over dynamic networks, as well as playback and recording of content with various devices. In this project, we have developed software for real-time MPEG-2 to MPEG-4 transcoding with a reduced bit-rate and spatio-temporal resolution. For surveillance applications, this enables MPEG-2 broadcast quality content to be received by a central service center and be distributed to remote clients over narrow-band networks. MPEG-4 is the preferred format such networks due to its coding efficiency and error robustness.

Background and Objectives: Recent advances in signal processing combined with an increase in network capacity are paving the way for users to enjoy services wherever they go and on a host of multimedia capable devices. Such devices include laptops and mobile handheld devices. Each of these terminals may support a variety of different formats. Furthermore, the access networks are often characterized by different bandwidth constraints, and the terminals themselves vary in display capabilities, processing power and memory capacity. Therefore, it is required to convert and deliver the content according to the network and terminal characteristics.

Technical Discussion: Our MPEG-2 to MPEG-4 transcoding software is able to achieve reduced bit-rates, spatial resolutions, and temporal resolutions. The transcoding is done in an efficient way such that multiple bitstreams can be transcoded with general-purpose processors. The brute force approach decodes the original bitstream to the spatial-domain, performs some intermediate processing, and then finally re-encodes to a new bitstream. Our proposed architectures simplify this process, while still maintaining the picture quality.

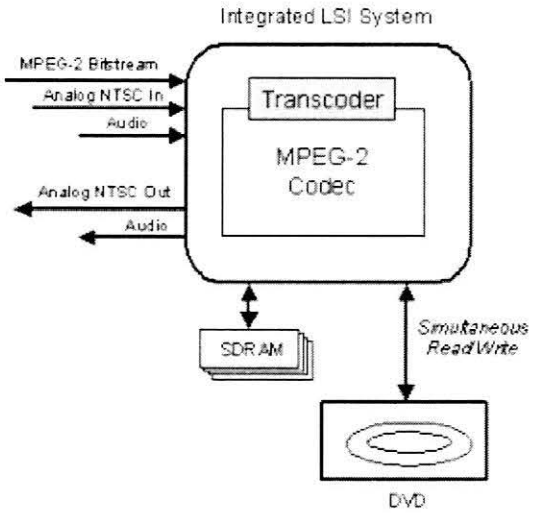
Collaboration: This project is done in collaboration with the Multimedia Information Coding & Transmission Technology Department at Johosoken and the Image Information Processing Department at Sentansoken. Princeton University has also contributed to this project.

Future Direction: Transcoding is a key technology for many communication systems. We continue to reduce the complexity of certain transcoding modules and work on improving the quality. Also, we seek new applications that this technology can be applied to.

Contact: Anthony Vetro
http://www.merl.com/projects/MPEG-transcoding/

Lab: MERL Murray Hill
Project Type: Advanced Development

MPEG Transcoding for DVD Recording



In general, the purpose of a transcoder is to convert compressed content, such as an MPEG (Moving Pictures Experts Group) bitstream, into a format that satisfies transport over dynamic networks, as well as playback and recording of content with various devices. In this project, we are developing a software model for an MPEG-2 transcoder that adapts the bit-rate and spatial resolution. The transcoding logic should be integrated with an MPEG-2 Codec to make an efficient LSI design. For DVD recording, the MPEG-2 broadcast contents may be MP@ML (Main Profile @ Main Level (Standard-Definition Television)) or MP@HL, (Main Profile @ High Level (High-Definition Television)) but the recorded DVD contents must be MP@ML. Also, the bits recorded to the disk have some limitations as well, and are usually variable bit-rate.

Background and Objectives: Digital television is now being broadcast around the world in the MPEG-2 format. Depending on the county, either MP@ML or MP@HL can be received. In the market today, you can find broadcast receivers, which are now being integrated into the newer digital television sets. Also, you can find a wide array of DVD players. The aim of this project is to be able to record the television broadcast onto the DVD medium in an efficient way. One important feature of the DVD recorder is simultaneous read/write.

Technical Discussion: The MPEG-2 Codec already includes an MPEG-2 MP@ML encoder to encode the analog NTSC input and an MPEG-2 MP@ML decoder to playback the contents from disk. The simplest way to record the broadcast content onto the DVD would be to employ a MPEG-2 MP@HL decoder with post-processing, then feed the resulting signal into the MPEG-2 MP@ML encoder. Instead of this complex solution, a more efficient transcoder is considered for integration with the existing MPEG-2 Codec.

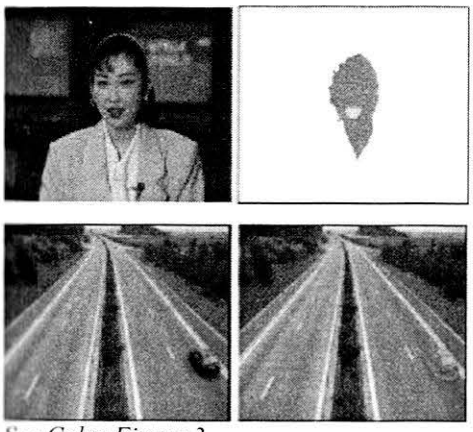
Collaboration: This project is done in collaboration with the System LSI Development Center, L-Ji-Se. Also, we discuss certain aspects of this project with the Multimedia Processor Department of Johosoken.

Future Direction: We continue to develop techniques for MPEG-2 based transcoding, and investigate the integrated LSI architecture.

Contact: Anthony Vetro, Hao-Song Kong
http://www.merl.com/projects/transcoding_dvd/

Lab: MERL Murray Hill
Project Type: Advanced Development

Video Object Segmentation



See Color Figure 3

Main purpose of video segmentation is to enable content-based representation by extracting objects of interest from a series of consecutive video frames. It is also a key to many robotic vision applications. Most vision based autonomous vehicles acquire information on their surroundings by analyzing video. Particularly, it is required for high-level image understanding and scene interpretation such as spotting and tracking of special events in surveillance video. For instance, pedestrian and highway traffic can be regularized using density evaluations obtained by segmenting people and vehicles. By object segmentation, speeding and suspicious moving cars, road obstacles, strange activities can be detected. Forbidden zones, parking lots, elevators can be monitored automatically. Gesture recognition as well as visual biometric extraction can be done for user interfaces. We developed a novel algorithm for automatic and reliable segmentation of moving objects in color video sequences and extraction of video object planes. A set of object descriptors are proposed to establish the relation between the different video objects hierarchically.

Background and Objectives: Our method has several advantages over the conventional techniques; it is automatic, computationally efficient, extracts object shape precisely, generates a multi-resolution object tree to expedite content analysis, and able to incorporate priori information.

Technical Discussion: After filtering the input video, markers are selected. Markers serve as the seeds of volumes. A volume is defined as the aggregation of video object planes of the same object in every frame of the sequence. Using the local color and texture characteristics, a volume is grown around each marker. The grown volumes are refined and motion trajectories are extracted. Self-descriptors for each volume, mutual-descriptors for a pair of volumes are computed from trajectories. These descriptors designed to capture motion, shape as well as spatial information of volumes. In the clustering stage, volumes are merged into objects by evaluating their descriptors. Iterative clustering is carried out until the motion similarity of merged objects becomes small. After clustering, an object tree that gives the video object planes for every possible number of objects is obtained.

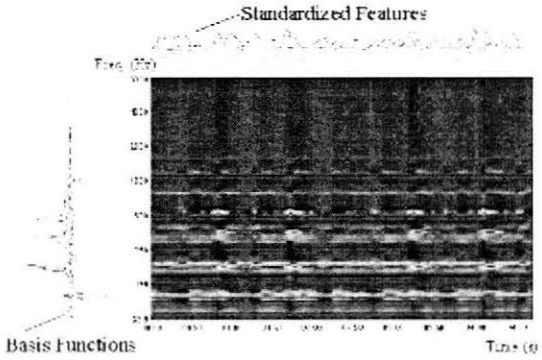
Collaboration: Joho-Soken and Sentan-Soken inside of MELCO, and with Polytechnic University, Brooklyn, NY.

Future Direction: Integrate object detection, identification, and tracking techniques into MELCO MPEG-4 encoder as well as video surveillance products such as traffic analyzer, intruder detector, and multi-camera surveillance system.

Contact: Fatih Porikli
http://www.merl.com/projects/video-segmentation/

Lab: MERL Murray Hill
Project Type: Research

MPEG-7 Sound Recognition



A spectrogram of an audio clip and its corresponding MPEG-7 features. Robust automatic sound classifiers may be built using such *signatures*.

MPEG-7 is the newest member of the family of industry standards for media technology published by the ISO. Released in July 2002, MPEG-7 standardizes media content indexing and retrieval for professional and consumer applications.

MERL CRL contributed several key technologies to the MPEG-7 audio standard supporting applications such as speaker identification, sound recognition, music classification and song similarity indexing.

MPEG-7 makes sounds, pictures and video as searchable as the text in Internet Web pages.

Background and Objectives: One of the major challenges in the design of sound recognition systems is selecting features and probability model parameters that are robust across a broad range of sound types. Robust systems should require no human intervention for feature extraction or model parameter estimation. To this end, we sought fully automatic methods for building recognition systems using training data.

Technical Discussion: The MPEG-7 standardized features for sound recognition consist of dimension-reduced spectral vectors obtained using a linear transformation of a spectrogram. Dimension reduction uses a MERL-patented technology, based on the singular value decomposition (SVD) and independent component analysis (ICA), to find a set of basis functions that maximize the information content of the features whilst minimizing their size. Such compact features are essential for efficient training of automatic classifiers and for robust performance.

Within the standard, these features are used with hidden Markov models to build robust automatic classifiers. HMM classifiers are represented within MPEG-7 using XML-based description schemes that enable interoperability and portability of models between different applications. The system successfully identifies sound events as diverse as speech, singing, environmental noises, animal sounds, musical instruments and music genres. Industry uses for this technology include remote audio monitoring, media archive searching and automatic music monitoring for broadcast facilities.

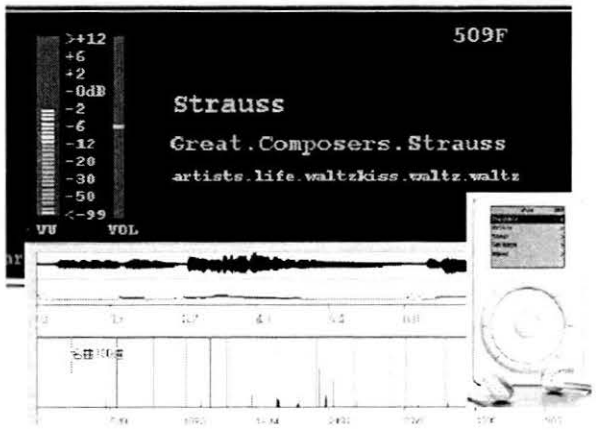
Collaboration: International Standards Organization. MERL MHL, MERL CSL

Future Direction: MPEG-7 music player, video browsing using sound information, other new business opportunities.

Contact: Michael Casey
http://www.merl.com/projects/MPEG7sound/

Lab: MERL Cambridge Research
Project Type: Advanced Development Project

MPEG-7 Music Player



By 2005, personal music devices will hold 10,000 or more songs. Such massive storage requires new interfaces for the user to be able to easily access content.

To address this problem, MERL CRL has been an active participant in MPEG-7 audio, a new international standard for media content indexing and retrieval. We have built a prototype music player that knows the sound of each song and uses song similarity indexing to find songs by mood, artist or style at the push of a single button.

Background and Objectives: Personal audio devices, such as MP3 music players, are extremely popular amongst music fans, especially those who like to download their content over the internet. Users can currently put up to 2000 songs on a single portable playback device, and will have access to up to 10,000 songs in the near future. The problem with these devices is that they lack any knowledge about the content and therefore offer no way for the user to navigate the enormous amount of information. MusicFinder addresses these problems by offering a query-by-example feature that finds lists of songs that are similar in style and mood to the current song.

Technical Discussion: MusicFinder is built upon MERL's contributions to the audio part of the MPEG-7 international standard for multimedia content description. Songs are indexed using a hidden Markov model (HMM) trained on many different types of music, much like a speech recognizer, but for musical signals. Histograms of the HMM output sequences are used as features for similarity matching on the musical signals. Musical structure (intro / verse / chorus) is extracted by clustering on the within-song similarity matrix providing audio summary information. Style or mood ordered playlists are generated by sorting the similarity scores between all pairs of songs in the database. MPEG-7 defines the methods and structures for extracting the indexes from audio content.

Collaboration: MusicFinder is currently being integrated into a larger framework called Media Finder, for music and video content retrieval, in collaboration with MERL CSL. Song and artist metadata indexing was developed in collaboration with the Massachusetts Institute of Technology (MIT). Many other labs within MELCO have been active within MPEG-7 including Johosoken, MERL MHL and VIL in Europe.

Future Direction: speech interface for music retrieval, user studies, new intellectual property.

Contact: Michael Casey
http://www.merl.com/projects/MusicFinder/

Lab: MERL Cambridge Research
Project Type: Research

Collaboration and Interaction

The bulk of MELCO's business is in manufacturing of devices, devices that are other than traditional desktop computers. These include projectors, white goods, antennas, power plant equipment, and cell phones. As these devices become "smarter," the best of them will at the same time become simpler to use. The goal is to build devices whose use is so natural that the human-computer interface "disappears." The device simply responds as it should.

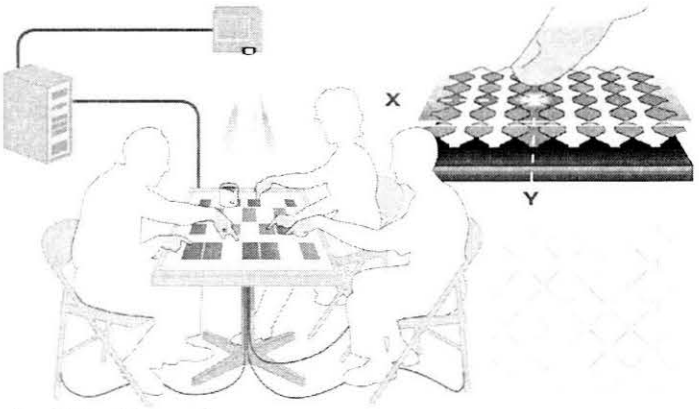
This technology area at MERL is exploring new smart devices and enhancements to existing devices. An example from MERL this year that received a lot of attention in the international press is interactive glassware (iGlassware). Glasses or other containers are instrumented to know how full they are and communicate their state to systems embedded in furniture or elsewhere in the environment. Another is Tangible Intermediaries, in which cameras are used to enhance digital functionality of actual objects in their natural environment.

A particular theme within the interaction area is technology for devices that aid human-human collaboration. The prime example is DiamondTouch, a multi-user touch technology that allows multiple users to interact with surfaces on tables or walls directly through touch. It is natural that more than one user be allowed to touch at the same time and that applications support the recognition of who is touching where. However, no prior existing technology was able to afford this. We are also developing the base technology for support of flexible viewing angles for information displayed on a circular table surface and manipulated and viewed by collaborating groups of people.

On more conventional desktop computers, MERL is exploring new methods for information visualization (Multi-Parametric Visualization) and intelligent training (Intelligent Agents for Operator Training and Task Guidance). These technologies are designed to support operations and new services growing from within the manufacturing sectors. For the future, MERL has also begun to explore issues related to interaction between humans and robots in Robot-Human Interaction for Hosting Activities.

Project Descriptions

DiamondTouch Hardware



See Color Figure 4

DiamondTouch is a simultaneous, multi-user, touch input device developed at MERL. Not only can it detect multiple, simultaneous touch events, but it can also identify who is touching where. This unique identification ability has made DiamondTouch a much sought after device in the university human-computer interface research community. We hope these efforts will ultimately lead to a strong commercial market for DiamondTouch systems. Work on DiamondTouch is proceeding on two fronts - improving the hardware, and creating a software development kit. This document discuss the hardware side of the effort.

Background and Objectives: DiamondTouch was initially created in 2001as an interface device for use in collaborative human guided search applications. We have since recognized the very general applicability of this technology, and are working towards creating commercially viable systems. By seeding select research university groups with prototype systems, we hope to create a base of compelling applications which will drive the market for these devices.

Technical Discussion: DiamondTouch uses an array of antennas embedded into a surface, with each antenna transmitting a unique signal. Each user has their own receiver, generally attached to their chair. When a person touches the surface, energy from nearby antennas is coupled through the user to their receiver. This is how the system determines who is touching where.

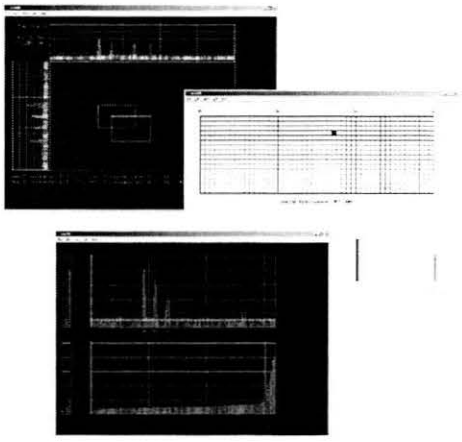
DiamondTouch hardware has gone through considerable refinement since the original concept prototype. The antenna array is now screen printed on plastic sheets which has allowed us to increase the size from 20cm x 20cm to 48cm x 80cm, decreasing costs while increasing performance. It is also possible to array these sheets to create much larger surfaces. Instead of a seperate box for the electronics, we have mounted the circuits inside the table, creating a self-contained unit. Currently, we are finishing our latest generation hardware which increases the scan rate, integrates 8 receviers on one board, and provides a USB interface.

Collaboration: This is a joint project between MERL-CS and MERL-CRL. MERL has an active collaboration with Johosoken and is seeking collaborations with other MELCO groups; we are also loaning DiamondTouch systems to a number of universities who will explore DiamondTouch as a collaborative input technology.

Future Direction: Current prototypes have non-standard aspect ratios. Future systems will have a 3:4 aspect ratio to better match currently available video projectors. We will also continue to improve the signal processing to increase scan rate and improve tolerance to noisy environments.

Contact: Paul Dietz, Darren Leigh **Lab:** MERL Cambridge Systems
http://www.merl.com/projects/DiamondTouch/ **Project Type:** Research

DiamondTouch Software Development Kit (SDK)



DiamondTouch technology is an input device that distinguishes between multiple simultaneous users and tracks multiple simultaneous touch points for each user. (See report on DiamondTouch for more details.) The DiamondTouch Software Development Kit (SDK) provides support for the development of Microsoft Windows and Linux applications that utilize DiamondTouch's capabilities to implement computer-supported collaboration and rich input modalities (such as gestures). When projected upon, the touch surface facilitates direct manipulation of user interface elements and provides a shared focus of attention for collaborating users. Possible applications include disaster-control command posts, power plant control rooms, business or technical meetings, and a variety of casual applications (musical instrument control, home coffee table, etc).

Background and Objectives: The DiamondTouch technology was developed as part of our Human-Computer Interaction effort at MERL Cambridge Systems. The SDK implements key features of the technology, provides a platform for further exploration of its possibilities and applications, and is the vehicle whereby we support our collaborators (internal and external).

Technical Discussion: The DiamondTouch hardware periodically produces frames of data indicating the proximity of the user's finger(s) to each antenna. The DiamondTouch Library (dtlib) reads these data frames from the device and affords access to the raw data and to various abstractions and interpretations of that data, such as the location of the touch point and the bounding box of the area touched. Other abstractions are possible and are the subject of ongoing research. A weighted interpolation algorithm increases the effective resolution to 2500 x 1500. Median filtering, hysteresis, and adaptive touch thresholding are used to improve robustness in the face of RF interference and other environmental variables. The SDK consists of dtlib (ANSI C), jdt (a Java interface layer), merldt (a Windows application providing mouse emulation, projector calibration, and various diagnostic displays), and a simple multi-user application example.

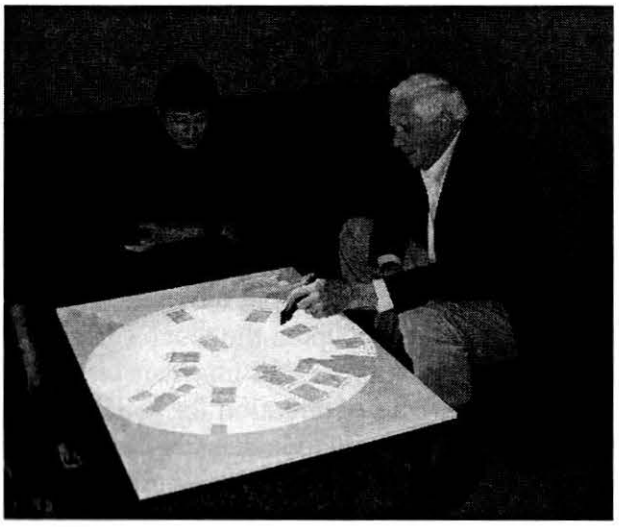
Collaboration: This is a joint project between MERL-CS and MERL-CRL. MERL has an active collaboration with Johosoken and is seeking collaborations with other MELCO groups; we are also loaning DiamondTouch systems to a number of universities who will explore DiamondTouch as a collaborative input technology.

Future Direction: We will provide ongoing support for our collaborators and plan to release a second generation of DiamondTouch with a native USB interface. We plan to investigate other input modalities (such as gestures), conduct more formal users studies and incorporate the results into future releases. Our main focus will be on adding value to existing or future MELCO products.

Contact: Kathy Ryall, Samuel Shipman
http://www.merl.com/projects/dtsdk/

Lab: MERL Cambridge Systems
Project Type: Research

DiamondSpin



DiamondSpin is a circular tabletop environment that preserves the simplicity and informality of around-the-table interaction, while at the same time provides a rich set of novel UI functions for interactive and collaborative document browsing, visualization, manipulation and navigation by small groups of people.

Background and Objectives: Tables are a familiar piece of furniture commonly found in homes, offices, cafés, design centers, show rooms, waiting areas, and entertainment centers. Tables provide a familiar and convenient physical setting for people to meet, chat, look over documents, and carry out tasks that require face-to-face collaboration. Digital documents, on the other hand, are commonly used only on desktop computers and handheld devices. Digital documents are much easier to share remotely than face-to-face, due to a lack of a physical media that contain the necessary computational support for face-to-face around the table applications. Our objectives are to research and study new HCI mechanisms which can enable tabletop applications. The output of this research will be a tabletop single display groupware toolkit which can be used to build new tabletop applications.

Technical Discussion: When multiple people gather around a table, there is no single directional viewing angle or orientation that is ideal for everyone present. We postulate that a polar-coordinate system, which is well suited to a circular display, can provide a continuous orientation among multiple people and can allow sharing of documents by rotating individual items or by rotating the entire display. The key research problems that the DiamonSpin architecture addresses stem from three unique characteristics of a user interface that is circular and is on a tabletop: (1) handling the polar location, orientation and deformation of documents on the table, (2) manipulating, displaying and refreshing large quantities of pixels from potentially many piled and overlapped documents, and (3) managing multi-user collaborative activities. So far, we have developed a transformation engine transforming polar coordinates into a standard transformation matrix for graphics context and input events (the T-Engine), and a multi-layered multiple depth asynchronous rendering engine (the MMR-Engine).

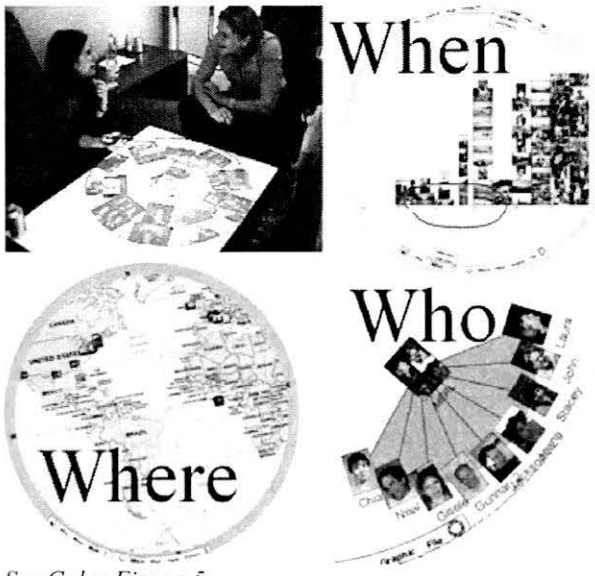
Collaboration: With Imperial College on user study of interaction patterns with opportunistic local information browsing on coffee tables. With MERL CSL on building multi-user DiamondSpin using the touch technology of DiamondTouch.

Future Direction: We are currently in the process of developing a variety of applications with the DiamondSpin architecture and expanding its multi-user functionality. We are planning more user studies to better understand interaction on tabletops.

Contact: Chia Shen, Neal Lesh
http://www.merl.com/projects/diamondspin/

Lab: MERL Cambridge Research
Project Type: Research

Personal Digital Historian (PDH)



See Color Figure 5

PDH is a new digital content user interface and management system. Unlike conventional desktop user interfaces, PDH is intended for multi-user collaborative applications on single display groupware. PDH enables casual and exploratory retrieval, interaction with and visualization of digital contents. PDH is built on top of our DiamondSpin circular tabletop environment. Our current project includes research in the areas of content annotation, retrieval and presentation, visualization of and user interaction with images, audio, video and data, as well as the study of how people collaboratively use the single display interface.

Background and Objectives: As part of people's daily life at work, on the go and at home, their computers, PDAs and digital cameras generate larger and larger amounts of digital contents. However, technologies that allow people to easily utilize this digital data in a face-to-face conversational or group setting are lagging far behind. Applications are limited by the user interface potentials of current desktop computers and handheld devices. The objective of the PDH project is to take a step beyond.

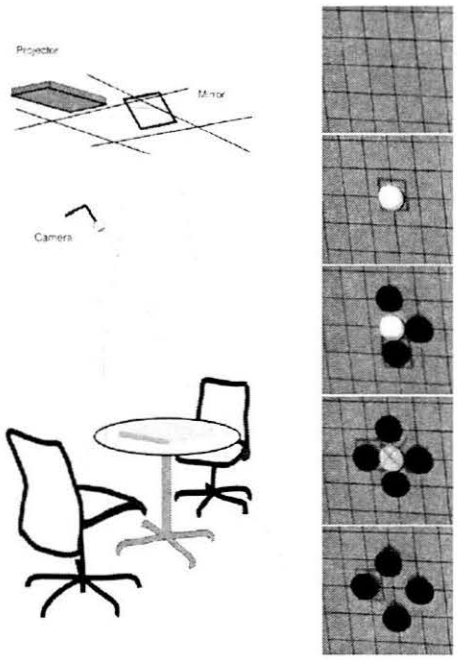
Technical Discussion: Creating a new type of interface requires addressing many issues. One of our primary focuses is on developing content organization and retrieval methods that are easy and understandable for the users, and can be used without distracting them from their conversation. Rather than the folder&file mechanisms used by conventional document systems, PDH organizes the contents along the four W's of storytelling (Who, When, Where, and What) and allows users to design new contexts for organizing their structures. A second issue we have focused on is affording casual and exploratory interaction with data by combining a multiplicity of user interaction mechanisms including in-place query and in-place pop-up menus, direct manipulation, natural visual query formulation with minimal menu-driven interaction and freeform strokes. Finally, in order to support the multi-threaded and non-linear progression of group conversation, PDH provides tools to help people navigate a conversation as well as their content.

Collaboration: We are developing semantics and mechanisms for simultaneous multi-user input and developing new application scenarios for PDH along with continued user studies.

Contact: Chia Shen, Neal Lesh
http://www.merl.com/projects/PDH/

Lab: MERL Cambridge Research
Project Type: Research

Tangible Intermediaries



The tangible intermediary is an idiom for the design of user interfaces that places the focus on augmenting existing physical objects with cast computation. We applied this notion as a design principle to the implementation of Golly, an augmented reality go board game. Golly augments a physical go board by fluidly integrating many of the computational enhancements previously only available with graphical user interfaces. After running an experiment to compare Golly to traditional graphical user interfaces we found that participants could interact with it more quickly. Additionally, we found that participants preferred Golly to traditional designs. It is our conjecture that the participants' preference for and increased performance using Golly are due to the ease with which users can learn interfaces based on the tangible intermediary idiom. Tangible Intermediaries represent an important direction both for user interface designers, and for projector manufacturers.

Background and Objectives: Tangible intermediaries represent a step in an interesting direction for ubiquitous computing. Tangible intermediaries are a way to layer computation onto existing objects: without replacing them, or adding LCD panels with GUI interfaces, or requiring a separate handheld for configuration. The user can interact with objects in their environment exactly the same way as they always have, but the computer can now sense those interactions, and can supply helpful feedback as an overlay, or take complementary actions.

Golly is a tangible intermediary for the game of go. Using vision techniques, an unmodified go board can be observed and gameplay recorded and analyzed. Making use of an LCD projector, the board can be augmented with information from the computer, for instance, by projecting a remote opponent's moves onto the board, displaying a clock to show how much time has elapsed during the user's move, or allowing users to back off several moves and explore variations.

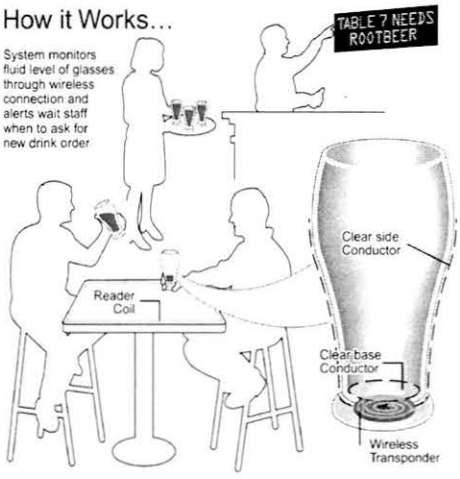
Technical Discussion: The playing surface is a common table viewed by a Firewire camera and illuminated with an LCD projector from overhead. The perceptual system is implemented primarily in Java, supported by assembly-coded image manipulation functions. The system runs on two Intel 1GHz workstations: one to process the firewire camera data, and a second to manage the display projector. The overall goal of the perceptual system is to understand the progress of the game well enough to support the interaction modalities discussed above.

Collaboration: Extending the Tangible Intermediary principle to other application domains.

Contact: Christopher R. Wren
http://www.merl.com/projects/tangible-intermediaries/

Lab: MERL Cambridge Research
Project Type: Research

iGlassware



Since restaurants often make much of their profits on drinks, it is critical for servers to offer refills in a timely fashion. We propose wireless liquid level sensing glassware to aid in this task. Specially instrumented glassware detects fluid levels via a high-resolution capacitance measurement. A coil embedded in the table inductively couples power to the glasses, and provides a path for data exchange. Our prototype glass uses a standard microprocessor and a small number of passive components, making it extremely inexpensive.

Background and Objectives: It is a common problem – you are in a bar or restaurant with your drink almost gone and you are desperately hoping that one of the staff will notice and offer you a refill. Sometimes they do, and sometimes they don't. If they don't, you leave a little less happy with your experience and are less likely to return, the waiter or waitress gets a lower tip, and the restaurant has lost the chance to sell you a drink. Meanwhile, thirsty customers may stand waiting at the door for lack of a table. Everyone loses. It is such a little thing; yet doing it right or wrong can easily make the difference between economic success or failure.

It is thus critical for servers to offer refills in a timely fashion. We propose wireless level sensing glassware to aid in this task. Ideally, instrumented glassware, or iGlassware, should have the following characteristics:

- Extremely inexpensive
- Washable by standard restaurant dishwashing equipment
- No maintenance issues (e.g. battery replacement)
- Familiar glassware appearance (no wires, not bulky, etc.)
- Support multiple glasses per table
- Globally unique IDs for each glass
- Able to recognize a glass of remaining ice as empty of fluid
- Reasonable measurement resolution.

By using a combination of RFID and capacitance sensing technologies, we are able to achieve these properties.

Technical Discussion: MERL will be working with Mitsubishi Electric business units for OEM product commercialization. In addition, MERL is interested in input from leading companies in the restaurant/hospitality industry in order to refine the technology and assess market potential.

Collaboration: MERL is now refining the prototype systems in preparation for field testing and potential product commercialization.

Contact: Paul Dietz, Darren Leigh, William Yerazunis
http://www.merl.com/projects/iGlassware/

Lab: MERL Cambridge Systems
Project Type: Research

Multi-Parametric Visualization

Movie Recommender with Data Mining and Visualization Techniques

Multi-Parametric Visualization is a set of methods and tools for visualization and querying of multidimensional information. It is designed to be easy to use and quick to yield insights into the data.

Background and Objectives: Information visualization tools, while an industry unto itself, have by and large remained difficult for nonspecialists to use. Particularly when the information to be visualized comes in the form of a relational table, it is rare for an executive or a designer to be able to quickly and flexibly explore the data and derive insight to aid decision making. The Multi-Parametric Visualization project at MERL has the objective of developing new techniques and tools that can improve engineering design, operations, and business decision-making at MELCO. In the longer run, we expect that these tools will also be included in service offerings to external customers in the power and utilities business among others.

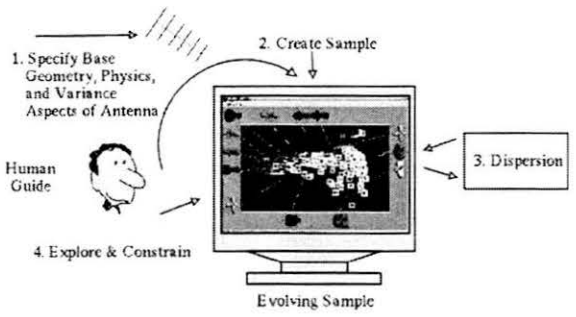
Technical Discussion: Multi-parametric visualization is based on a set of parallel visualizations called bargrams. There is typically one bargram for each column of information in a table. Bargrams are similar to histograms, revealing distribution of counts across an ordered set of "bins" or categories. However, the bargram saves visual real estate and offers additional visualization features by "tipping over" the histogram to create a simple one dimensional visualization for each attribute. Users can quickly create queries and preview their results by selecting subsets of attribute values. For research purposes, MERL has licensed software from Verizon, created by members of the current staff at CSL when they were at Verizon Laboratories. This has allowed the project to quickly prototype new application domains and create requirements for future work. MERL has also developed a new implementation of MPV for the Human-Guided Antenna Design project that allows designers to quickly evaluate and winnow down a large set of generated antenna designs based on their collective features.

Collaboration: Real-time Recommender Systems and Human-Guided Antenna Design projects at MERL and communications and power systems business units at MELCO.

Future Direction: The project expects to develop a comprehensive set of middleware that will allow for rapid development of multi-parametric visualization components for a variety of MELCO applications.

Contact: David Wong, Frederick J. Igo, Jr., Kent Wittenburg **Lab:** MERL Cambridge Systems
http://www.merl.com/projects/mpv/ **Project Type:** Research

Human-Guided Antenna Design



Optimization-based approaches to antenna design have enjoyed limited success. The task is often computationally intractable. Moreover, it is also often difficult to capture all relevant design issues and tradeoffs in a single mathematical objective function. Therefore, human experts typically specify and refine antenna designs by hand, using computers only to evaluate their candidate designs by simulation. In this project we propose a middle ground between this traditional approach and fully automatic optimization -- a human-guided interactive system.

Background and Objectives: The idea of using computer-based optimization for design tasks has been applied to many problems, including antenna design. However, this idea does not always work well: the optimization problems are often intractable; and it often proves impossible to consider all relevant design criteria in the optimization process. In this project we propose that the computer be used differently. Instead of having the computer search for a single optimal design, we program it to intelligently sample the large space of possible antenna designs, subject to user-supplied constraints. The task of choosing a final design from the computer-generated sampling is left to the human user, who can apply experience and judgment to recognize and then refine the most useful antenna design.

Technical Discussion: At the heart of our approach, the computer generates a sample set of candidates (called a population) of possible antenna designs and presents them to the user. The person's role is to help guide the computer in generating a desirable population, and ultimately to select an appropriate antenna.

The two key components in our system are dispersion and visualization. Dispersion is the process by which we generate a representative sample of designs; it is an intelligent sampling process. A key requirement for dispersion is a mathematical function that quantifies the difference between two antenna designs. This difference metric is usually based on the performance characteristics of an antenna. Visualization is the process by which a person can examine the current population to examine trade-offs in the design space (e.g., cost and gain) and to constrain future searches. We are currently using MPV (multi-parametric visualization, see /projects/mpv) as our visualization component. The design process is an iterative one, alternating between dispersion and visualization.

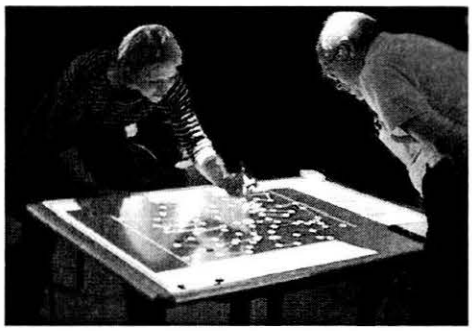
Collaboration: This project is a joint effort between MERL Cambridge Systems and MERL Cambridge Research in collaboration with Johosoken and with sponsorship from Denshihon.

Future Direction: We have completed a proof-of-concept implementation and will be testing its design on variety of antenna problems to ascertain its strengths and weaknesses. Future work will involve incorporating a more sophisticated simulator into the system. We are also investigating generalizing this approach in the Human-Guided Search and Visualization initiative.

Contact: Kathy Ryall
http://www.merl.com/projects/antenna/

Lab: MERL Cambridge Systems
Project Type: Research

Human-Guided Search



Interactive, or human-in-the-loop, optimization systems leverage people's abilities in areas in which they outperform computers, such as visual and strategic thinking. Users can steer interactive optimization systems towards solutions which satisfy real-world constraints. Furthermore, people can better understand, justify, and modify solutions if they have participated in their construction.

We have developed the Human-Guided Search (HuGS) framework and Java toolkit for rapidly developing interactive optimization systems. The framework and code include visual metaphors for focusing and constraining optimization algorithms. The user can select from different algorithms, including a human-guidable version of a powerful heuristic, called tabu search. We have developed a wide variety of applications with the HuGS toolkit, including interactive systems to solve scheduling, vehicle routing, layout, and protein-folding problems.

Background and Objectives: This work represents a multi-person, multi-year, ongoing effort to develop experience, techniques, and generic software to produce interactive optimization systems. Our goal is to overcome many limitations of almost all current optimization systems. Current optimization systems typically solve an oversimplified formulation of the real-world problem and produce solutions which are difficult for users to understand or trust. In contrast, interactive optimization allows users to explore many possible solutions in order to better understand the tradeoffs between possible solutions and then choose a solution based on their rich understanding of the domain.

Technical Discussion: Our framework provides the user a greater degree of control than previous interactive optimization approaches. Users can manually modify solutions, backtrack to previous solutions, and invoke, monitor, and halt a variety of search algorithms. More significantly, users can constrain and focus the search with a visual metaphor that we have found effective on a wide variety of problems. Our experiments have shown that human guidance can improve the performance of the exhaustive search algorithm on the capacitated-vehicle-routing-with-time-windows problem to the point where the interactive algorithm is competitive with the best previously reported algorithms. Further experiments on other problems has shown that 10 minutes of guided tabu search is comparable to, on average, 70 minutes of unguided tabu search.

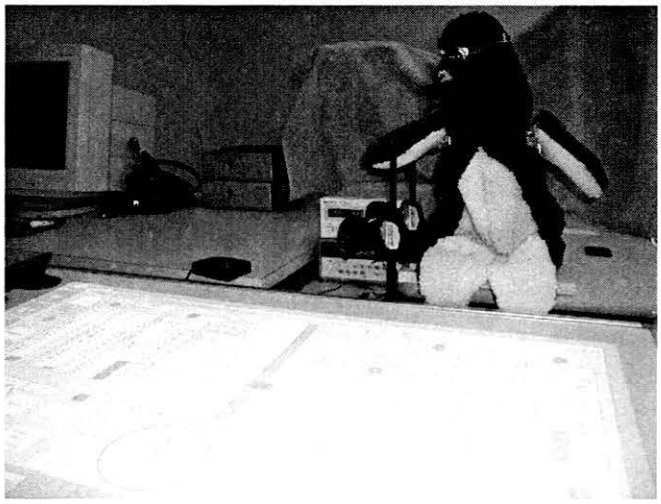
Collaboration: We have collaborated with researchers at Sentansoken on a real-world vehicle routing problem. We are actively collaborating with researchers at Harvard University, Vienna University of Technology, and McGill University.

Future Direction: We are continuing to refine our scheduling and protein-folding applications. We are developing a new application for the commercially-important polygon-packing problem. We plan to parallelize our algorithms and explore multi-user interactive optimization in both co-present and remote conditions.

Contact: Neal Lesh, Joe Marks
http://www.merl.com/projects/HuGS/

Lab: MERL Cambridge Research
Project Type: Research

Robot-Human Interaction for Hosting Activities



We have developed speech and gesture capabilities for a non mobile robot that uses COLLAGEN middleware for conversational hosting tasks.

Background and Objectives: Our objective is to develop spoken-language conversational systems which not only speak to humans, but use physical gestures and movement to guide them. A situation in which this kind of interaction is particularly appropriate we call "hosting." In a hosting activity, the "host" (which may be a human or a robot), provides the "visitor" (typically a human) with various kinds of information, entertainment, education, etc., related to their shared physical environment. Examples of human roles that include hosting activities are docents, receptionists, and real-estate agents.

Technical Discussion: We developed a prototype hosting robot, embodied as a penguin, using the COLLAGEN collaborative agent middleware to implement its conversational component. The robot has the ability to point (with its beak) at objects in its physical space, such as the buttons and dials on a simulated engine control panel projected on a nearby table top. We used the MIT Media Lab's BEAT system to compute and coordinate the robot's wing and beak gestures in coordination with its spoken utterances.

In our demonstration scenario, the robot acted as am embodied tutor for a gas turbine engine training system, which was also developed at MERL using COLLAGEN. The body motions of the robot served to helped users locate the appropriate buttons and dials they needed to manipulate and attend to as part of learning how to operate the engine.

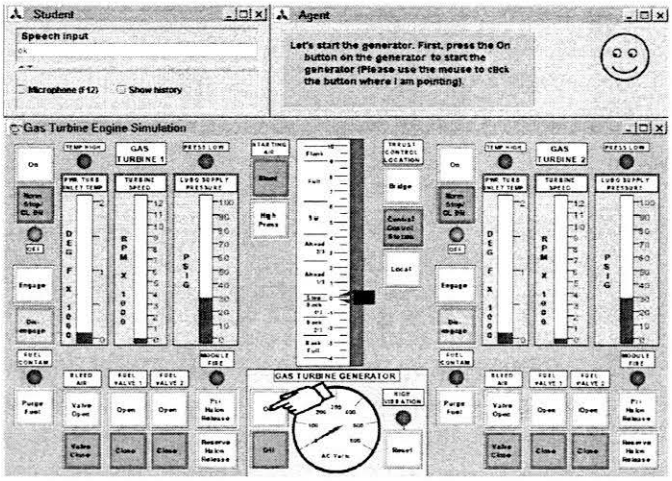
Collaboration: We collaborated with the MIT Media Lab in the use of the BEAT gesture generation toolkit, which they developed.

Future Direction: We plan to add machine vision and auditory processing capabilities to the robot in order for it to be able to physically locate the human with whom it is interacting. We also plan to experiment with other hosting environments.

Contact: Candace Sidner, Neal Lesh, Charles Rich
http://www.merl.com/projects/hosting/

Lab: MERL Cambridge Research
Project Type: Research

Intelligent Agents for Operator Training and Task Guidance



Intelligent software agents can help the operators of complex industrial equipment both with training and in the performance of their tasks. During training, the operators practice their tasks in a simulated environment. The intelligent agent guides operators through the steps of complex procedures, giving positive and negative feedback on their actions, and adapting to their skill level. We have developed generic software for building such software agents and have demonstrated it by building an agent for training operators of a simulated gas turbine engine.

Background and Objectives: This work extends the COLLAGEN middleware for building collaborative agents. Our goals are to enhance existing training systems by adding more sophisticated tutorial strategies and also to build new training and task guidance systems. Potential application areas include supervisory control and data acquisition (SCADA) systems, such as industrial plant control centers, equipment maintenance tasks, such as elevator repair, and the use of complex software interfaces, such as computer-aided design tools.

Technical Discussion: Our approach to operator training is based on "learning by doing." The software agent guides the operator through a sequence of example scenarios that incrementally expose the operator to the full complexity of the task to be learned. During the training process, the agent maintains a model of the operator's proficiency in each part of the task, so it can appropriately introduce new subtasks as well as give the operator the opportunity to practice previously taught knowledge.

Using COLLAGEN as our implementation base gives the training system developer two major advantages. First, we have an application-independent architecture in which pedagogical strategies are encoded in an application-independent manner; this allows the developer to reuse a large amount of code when creating a new training agent. Second, the same task model created to support a training agent can be reused to produce an agent which will act as an advisor or intelligent assistant to an operator during the actual performance of his job, if desired.

Collaboration: We are currently working with Mitsubishi Electric's Advanced Technology R&D Center, and Energy and Industrial Systems Center to produce a prototype agent to train thermal power plant operators using a software simulator.

We are also collaborating with the University of Southern California, Information Sciences Institute and the MITRE Corporation on embedded training applications in general.

Future Direction: We are planning to expand the range of pedagogical strategies used by our training agents, improve their robustness, and perform user evaluations.

Contact: Neal Lesh
http://www.merl.com/projects/training/

Lab: MERL Cambridge Research
Project Type: Research

Computer Vision

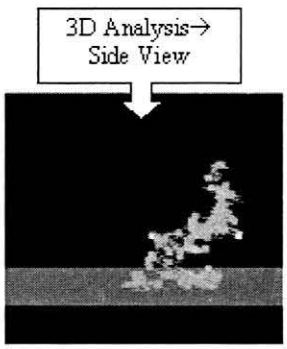
Computer Vision is the analysis of images to extract information about the world, the same functions that the human visual system provides (although perhaps accomplished through different mechanisms.) As sensor and computer hardware drops in cost, these visual functions can become features in a wide range of products where they provide automatic, fast, convenient, and precise alternatives for tasks that were previously manual.

Much of the Computer Vision research at MERL is concerned with 3D; it is based on estimating 3D location, tracking objects in 3D, modeling 3D object surfaces, etc. These processes make use of simultaneous images from multiple cameras or perhaps multiple images from a single camera over time. These functions are analogous to human stereovision. In many cases the 3D model of the world is used as input to make a process automatic (e.g. registration among multiple projectors) or to provide a user interface for human control (e.g. fast detection/location of a finger touching a screen.)

Additional MERL research seeks to extract other forms of information from images (e.g. to detect and recognize faces) or to create a new image from the original (e.g. a new image with higher resolution, or from a different viewpoint.) MERL has a special program called Computer Human Observation (CHO) which is developing and tailoring the computer vision techniques mentioned above to provide better security, safety, and system responsiveness.

Project Descriptions

Computer Human Observation (CHO)



See Color Figure 6

CHO is a three-year project aimed at devising new computer vision tools for automatic detection and recognition of people and analysis of human behaviors. In addition, we will provide a development platform for those tools that will allow easy access for MELCO laboratories and business units. We would like to enable robust computer analysis of video images for answering questions like: Where are the people? How many are there? What are their characteristics? Do they need anything? Can any be recognized?

Background and Objectives: Many of MELCO's businesses can benefit from more precise information about people. In surveillance applications, the staff required to monitor the systems already in place now dominates costs. Software that automates tasks that are currently manual will be quite attractive and the chief differentiator among systems. In elevator and building applications, effective detection of humans and their actions can promote safety and allow systems to anticipate how best to serve their users (e.g. are there any special needs or emergencies detected?) In retail settings, better understanding of how people are responding to the environment you create can help optimize that setting to make it most effective. We believe that we can create new business opportunities for MELCO by developing and applying better algorithms.

However, new technology is not sufficient. Another CHO objective aims to support transfer of that technology into MELCO products by making it available in a well-engineered software platform that allows our colleagues to easily adopt code modules for incorporation into products, prototypes, and/or experimental systems.

Technical Discussion: Many of the problems that push the envelope in computer vision are "real world" tasks like counting the crowd outside an elevator or identifying the valid maintenance crew for access to the jetway. Cost and speed are part of the envelope—a slow, expensive solution may be a non-starter. Therefore we strive to invent, develop, and test in the context of real applications. This has led us to state-of-the art algorithms for face detection and recognition, fingerprint verification, and stereo-based three dimensional object analysis.

Collaboration: The Vision Systems Group in Sentansoken and the Audio-Visual Content Analysis Group at MHL are also contributing to the CHO project.

Future Direction: With a strong library of image analysis tools, in the future we aim to combine tools to solve specific safety, security, and video analysis problems.

Contact: Jay Thornton
http://www.merl.com/projects/CHO/

Lab: MERL Cambridge Systems
Project Type: Initial Investigation

Biometrics Using Stereo Vision

The goal of this project is to extract appearance descriptors for people, based on whole-body appearance. A target application is to observe a person who appears in front of one surveillance camera (say at an entrance to an airport), then to reacquire that person if they reappear later in front of another camera (say at a check-in counter). This type of reacquisition is essential to analyze the movement of people through extended environments such as airports, factories and buildings.

Background and Objectives: There is no doubt that faces are a key cue for reliable passive recognition of individuals. But good quality images of faces may not be available in many surveillance situations - the camera resolution may be low, or an elevated camera may be capturing only oblique views of the face if the subject's head is down. There is a range of other cues - including height, body shape, color of clothing - which are weaker than face recognition but which can be reliably captured using typical deployments for surveillance cameras (e.g. cameras above doorways, or in the top corner of a room). The goal of the project is to investigate the measurement and use of these cues, with a focus on 3D measurements.

Technical Discussion: The basic technology uses stereo for computing 3D data for a scene. There are two key components to the work (a) segmenting individuals in a group of people, and (b) characterizing the appearance of an individual. The work on (a) has focused so far on reliable detection of edge chains around an individual's silhouette (by using stereo disparity to link edges with similar depth) as a precursor to recognition of silhouettes. The work on (b) has been on measurements of height and body-shape, with initial work on gait-measurements such as stride length and sideways motion of the head during walking.

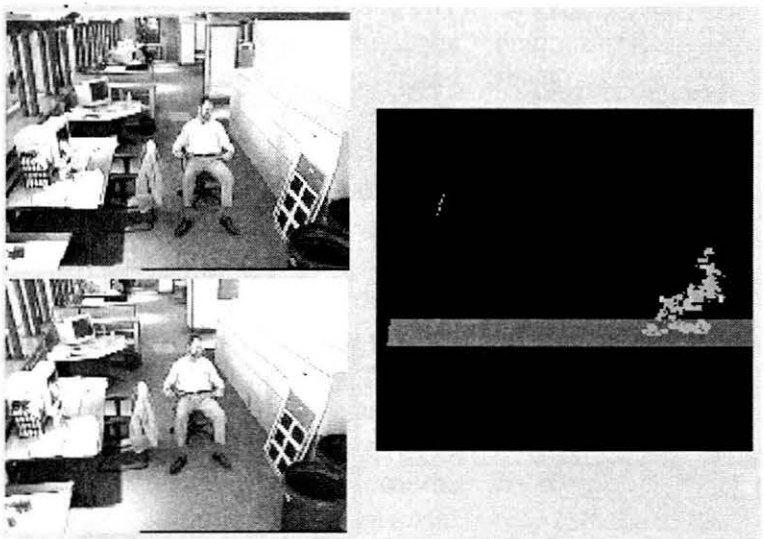
Collaboration: This work is part of the CHO SK.

Future Direction: This technology is of general use for surveillance applications.

Contact: Paul Beardsley, Emmanuelle Bourrat
http://www.merl.com/projects/biostereo/

Lab: MERL Headquarters
Project Type: Research

Wheelchair Detection Using Stereo Vision



The goal of this project is to automatically discriminate between ambulatory people and wheelchair users waiting at an elevator, using computer vision. A trinocular stereo system is used to (a) detect newly-appeared objects in front of the elevator, (b) recover 3D shape of the objects, (c) use 3D shape to identify wheelchair users.

Background and Objectives: Many MELCO elevators already provide modified service when a wheelchair is present. But the presence of the wheelchair is signaled by manually pressing a special button - we wish to replace this button with automatic wheelchair detection.

Technical Discussion: A trinocular stereo rig is mounted at the top of an elevator frame looking toward the exterior environment. From a single image, there are no strong visual cues to discriminate between an ambulatory person and a wheelchair user (the wheelchair itself is only partly visible given this type of frontal view). But there are valuable cues with multiple images, and we use stereo for measurements of height and 3D shape, plus temporal analysis to detect the foot motion of a walking person (versus the rigid feet of a wheelchair user).

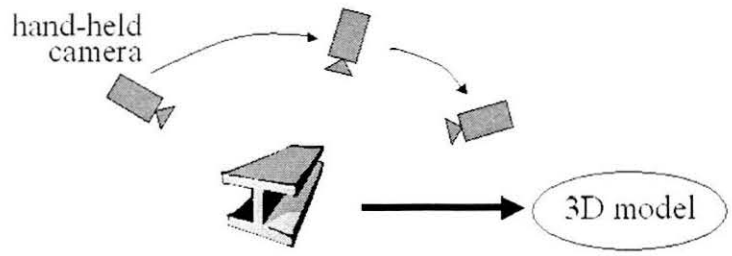
Collaboration: The software has been transferred to Sentansoken. The results will be presented to Inaden.

Future Direction: The current system works with single individuals in the scene. New work is on segmenting individuals from a group of people.

Contact: Paul Beardsley, Emmanuelle Bourrat
http://www.merl.com/projects/wheelchair/

Lab: MERL Cambridge Research
Project Type: Research

Hand-Held 3D Scanning Using Computer Vision



The goal of this project is a hand-held camera which is moved freely around an object in order to compute a 3D model of that object. The device actually consists of multiple cameras, typically two cameras. The main research problem being addressed is the robust recovery of camera motion during unconstrained motion. The initial work is on scanning objects with dimensions up to about 0.5m. Future work will target larger objects, for example automobile-sized objects.

Background and Objectives: Existing 3D scanner systems which have the most reliable performance include (a) turntable-based systems in which a fixed camera or cameras scan an object rotating on a turntable, and (b) cameras with attached motion-sensors (such as ultrasound). Problems are expense and the inflexibility of a fixed installation. The objective of this work is to do 3D scanning with a hand-held two-camera device. The advantages are cheapness, and ease-of-use in an arbitrary setting, say around a factory, or in an office, or around the home.

Technical Discussion: The hand-held scanner has two radically-directed cameras, one of which is used purely to determine the motion of the device, and one of which is used purely to acquire images of the object of interest. Dividing the vision functionality between different cameras in this way, and making use of distinctive points in the environment, enables robust and accurate recovery of camera motion regardless of the type of object being scanned. This contrasts with a vision system which attempts to recover camera motion from images of the object being scanned - in which case recovery of camera motion can be adversely affected by conditions such as homogeneity, transparency, or specularity of the object.

The complete scanning system involves, in addition to recovery of camera motion, two other main components (a) foreground/background segmentation for arbitrary backgrounds, which is the subject of current work, (b) model-building (which has been the research focus of the MERL 3D Images project).

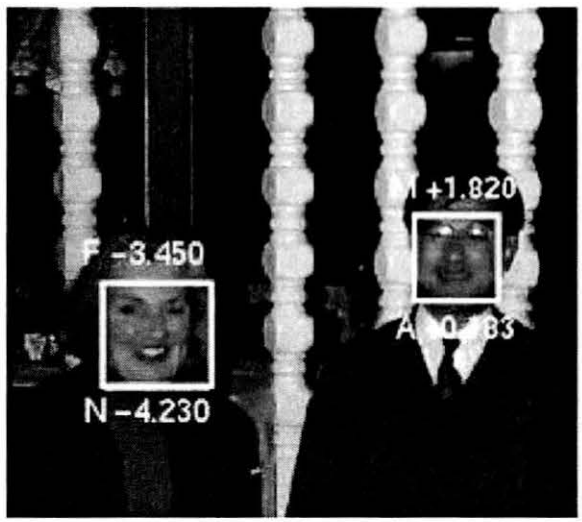
Collaboration: Within MELCO, we are in discussion with SentanSoken about possible applications.

Future Direction: Future work will be on scanning larger objects e.g. people- sized, or automobile-sized. We are also applying the same technology to a hand-held projector, looking at recovery of projector motion to enable stable/correct projection of augmented reality information from a hand-held projector.

Contact: Paul Beardsley
http://www.merl.com/projects/handheld3d/

Lab: MERL Cambridge Research
Project Type: Research

A Unified Approach for Face Detection, Face Analysis, and Surveillance



Automatic face and person detection is a critical component in the new domain of computer human observation and computer human interaction (HCI). There are many examples including: user-interfaces that can detect the presence and number of users; teleconference systems can automatically devote additional bandwidth to participant's faces; and video security systems can record facial images of individuals after unauthorized entry.

This is an extension of our previous work on fast face detection. Our new results include a system that can determine the gender (male/female) and race (asian / non-asian) of the face. Other new results include a detector for people in low-resolution surveillance video.

Background and Objectives: In the past many other face detection approaches have been proposed – based on neural networks, support vector machines, and other types of machine learning. No existing approach has achieved 'real-time' processing. Our goal was to develop a highly reliable yet extremely efficient face detectors and classifiers. Underlying this detector is a new machine learning approach that can be used for a variety of related tasks.

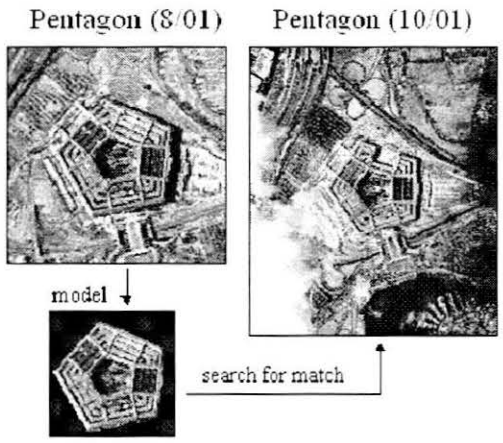
Technical Discussion: There are three main contributions of our object detection framework. First: a new image representation called an Integral Image that allows for very fast feature evaluation. Second: a method for constructing a classifier by selecting a small number of important features using AdaBoost. In order to ensure fast classification, the learning process must exclude a large majority of the available features, and focus on a small set of critical features. Third: a method for combining successively more complex classifiers in a cascade structure which dramatically increases the speed of the detector by focusing attention on promising regions of the image. The notion behind focus of attention approaches is that it is often possible to rapidly determine where in an image an object might occur. More complex processing is reserved only for these promising regions.

Collaboration: We are currently extending this approach to face recognition and to the detection of cars in video roadway video.

Contact: Paul Viola, Michael Jones
http://www.merl.com/projects/FaceDetection/

Lab: MERL Cambridge Research
Project Type: Research

Factorized Local Appearance Models



See Color Figure 7

We propose a novel scheme for image-based object detection and localization by modeling the joint distribution of k-tuple salient point feature vectors which are factorized component wise after an independent component analysis (ICA). Furthermore, we use a distance-sensitive histograming technique for capturing spatial dependencies which enable us to model non-rigid objects as well as distortions caused by articulation.

Background and Objectives: For appearance based object modeling in images, the choice of method is usually a trade-off determined by the nature of the application or the availability of computational resources. Existing object representation schemes provide models either for global features or for local features and their spatial relationships. With increased complexity, the latter provides higher modeling power and accuracy. Among various local appearance and structure models, there are those that assume rigidity of appearance and viewing angle, thus adopting more explicit models while others employ stochastic models and use probabilistic distance/matching metrics. Our objective is to model the high-order dependencies of local image structure by estimating the complete joint distribution of multiple salient point feature vectors using a density factorization approach.

Technical Discussion: We construct a probabilistic appearance model with an emphasis on the representation of non-rigid and approximate local image structures. We use joint histograms on k-tuples (k salient points) to enhance the modeling power for local dependency, while reducing the complexity by histogram factorization along the feature components. Although, the gain in modeling power of joint densities can increase the computational complexity, we propose histogram factorization based on independent component analysis to reduce the dimensionality dramatically, thus reducing the computation to a level that can be easily handled by today's personal computers. For modeling local structures, we use a distance-sensitive histograming technique. A clear advantage of the proposed method is the flexibility in modeling spatial relationships. Experiments have yielded promising results on robust object localization in cluttered scenes as well as image retrieval. Most recently we have adopted parametric models using mixture of Gaussians with resulting enhancements in performance.

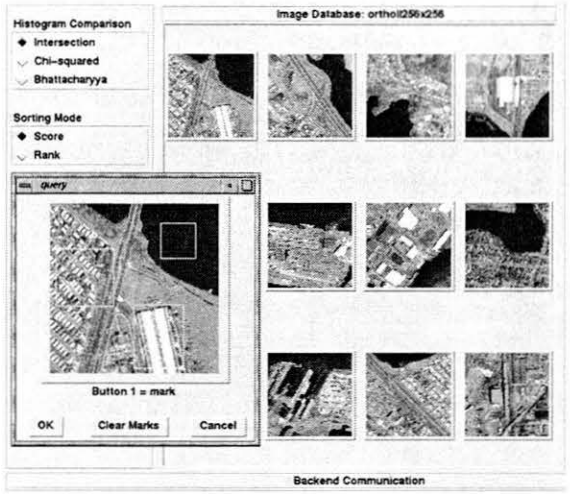
Collaboration: This project is joint collaboration Xiang Zhou and Thomas S. Huang (UIUC) and David Guillamet (University of Barcelona)

Future Direction: In the future, we plan to explore a more explicit way to incorporate spatial adjacency into the factorized local appearance model via graph matching. Also we are especially interested in applications of this technology to satellite image and photo-reconnaissance as shown in the example image above.

Contact: Baback Moghaddam
http://www.merl.com/projects/flam/

Lab: MERL Cambridge Research
Project Type: Research

Image Retrieval with Multiple Regions-of-Interest



With the proliferation of multimedia, the web and digital imaging, there now exists a high demand for intelligent tools for image management, most importantly indexing, search and retrieval, commonly referred to as QBIC or "query-by-image content." The goal of this project has been to develop a new image retrieval system based on the principle that it is the user who is most qualified to specify the "content" in an image and not the computer. The user is asked to provide salient ROIs or "regions-of-interest" and specify their spatial arrangements in the query image. This technique leads to a much more powerful image retrieval tool.

Background and Objectives: Most current "query-by-image-content" database indexing and retrieval systems rely on global image characteristics such as color and texture histograms. While these simple descriptors are fast and often do succeed in capturing a vague essence of the user's query, they more often fail due to the lack of higher level knowledge about what exactly was of interest to the user in the query image. The goal of this project was to develop and test a new technique using local image representations, grouping them into multiple user-specified "regions-of-interest" while preserving their relative spatial relationships in order to build a more powerful search engine for various applications of image database retrieval.

Technical Discussion: In our system we subdivide the image into an array of 16-by-16 pixel blocks each of which contains the following feature representations: a joint color histogram in LUV color space and joint 3D histogram consisting of the edge magnitude, Laplacian and dominant edge orientation, computed at two octave scales. These non-parametric densities represent local color and texture and due to the additive property of histograms can be easily combined to form bigger image blocks. When the user specifies a region of interest, its underlying blocks are "pooled" to represent a "meta-block" to be searched for in the database. Multiple regions are likewise searched and the intersection of the best matches determines the final similarity ranking of images in the database. In addition, the user can specify whether multiple selected regions should maintain their respective spatial arrangement.

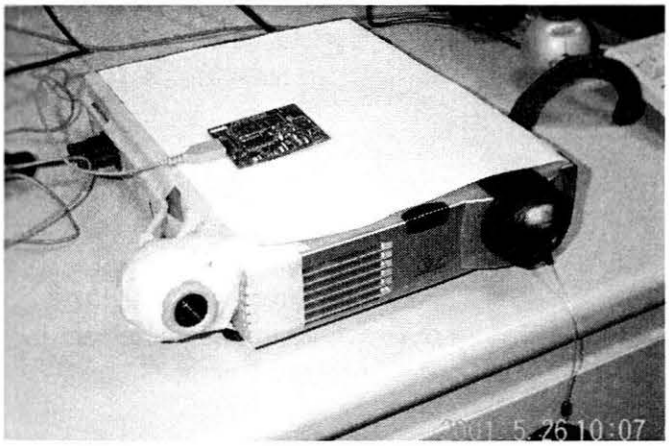
Collaboration: students from New York University and Carnegie Mellon University aided this project.

Future Direction: Currently the search for ROIs is computationally intensive and pruning strategies should be implemented in order to avoid searching the entire database for a "meta-block" query. Branch and bound techniques are prime candidates for speed-up.

Contact: Baback Moghaddam
http://www.merl.com/projects/idbr/

Lab: MERL Cambridge Research
Project Type: Research

Easy Calibration of a Projector



We have developed an easy-to-use technique to calibrate a projector. The same technique can also be used to calibrate a projector-camera stereo pair. The technique is useful because the calibration is achieved with a black planar surface (e.g. a wall or screen). No complex patterns or apparatus with Euclidean information are required.

Background and Objectives: Several techniques are currently available to calibrate a camera or a stereo pair involving two cameras. For example, camera internal parameters (such as focal length and principal point) can be calculated by viewing a planar checkerboard pattern at two or more different orientations. Relative pose (i.e. external parameters such as translation and rotation) between two cameras in a stereo pair can be calculated if both cameras can see the same checkerboard pattern.

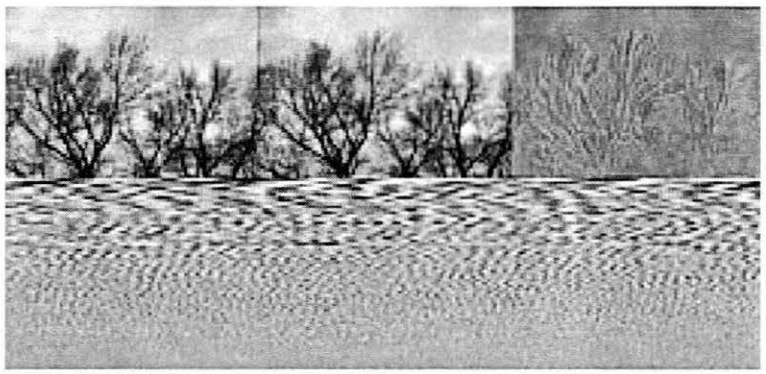
However, to date, there have been no simple techniques to calculate internal parameters of a projector. There are two reasons for the difficulty. First, since a projector cannot "see" a pattern, a feedback sensor (e.g. a camera) must be used. Second, for a projector the principal point is intentionally shifted (vertically) from the image center to allow an off-axis projection cone. Thus, the traditional assumption about the principal point (being close to image center) is not valid.

Technical Discussion: We use a rigidly attached camera. The camera is calibrated separately (by observing a printed checkerboard at two or more different orientations). To calibrate the projector, the same attached camera observes checkerboard pattern projected on a blank planar surface, at two or more different orientations with respect to the projector-camera stereo pair. The white planar surface could be a wall or fixed screen, and the rigid stereo pair is rotated to create the different relative orientations.

Using two or more homographies between projector and camera, we first compute the parameters up to a projective scale. We then upgrade the parameters in projective coordinates to Euclidean coordinates by assuming a fixed aspect ratio.

Contact: Ramesh Raskar, Paul Beardsley, Jeroen van Baar **Lab:** MERL Cambridge Research
http://www.merl.com/projects/ProjectorCalib/ **Project Type:** Research

System Identification for Video Texture



The top images at left are taken from a synthetic video of trees blowing in the wind. The rightmost image shows the difference between the two synthesized images, highlighting the sway of branches and the roll of the clouds. The video has completely realistic visual and dynamic texture, demonstrating the correctness of our novel system identification algorithm for high-dimensional sources such as video. The bottom image shows a temporal coding of the motion; each horizontal stripe indicates the independent swaying of a branch or motion of a segment of cloud. Only the first 1/3 of the motion is real; the rest is synthesized.

Background and Objectives: System identification is the problem of finding a model that fits data well enough to generate synthetic datapoints that are indistinguishable from real ones. We consider the problem of identifying a system that can synthesize realistic-looking video given a short training sequence. We conjecture that a linear dynamical system (a.k.a. Kalman filter) will suffice for a large range of natural phenomena, for example videos of rain, ocean waves, plants swaying in the breeze, the sky, facial motion, waterfalls, etc. The goal is to identify the system and its dimensionality, then solve for its parameters. The problem is that video images are extremely high-dimensional sources, while the phenomenon being modeled is probably generated by a low-dimensional system.

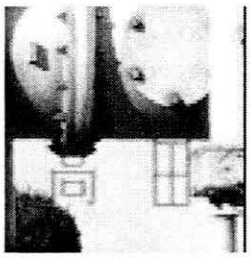
Technical Discussion: A linear dynamic system that generates observations Y=[y(1),...,y(T)] is defined y(t)=C*x(t)+D*e(t)+v(t); x(t)=A*x(t)+B*e(t)+u(t) where A is the system evolution matrix, B is the input matrix, C is the observation generating matrix, D is the feed-through matrix, X=[x1...xT] is the "hidden" state, and u and v are noise sources. We treat system identification (solving for A,B,C,D and the covariance of u and v) as a least-squares factorization problem and solve it explicitly. While not optimized for simulation and control, the solution does have excellent synthesis properties and is highly suitable for video and audio synthesis.

Collaboration: This work is directly applicable in any industrial setting where Kalman filters are used. We are looking at extensions to higher-order statistics and applications to animation control.

Contact: Matthew Brand
http://www.merl.com/projects/SystemID/

Lab: MERL Cambridge Research
Project Type: Research

Fast super-resolution method


Original High Res Image


Bicubic Interpolation


Super-Res on Luminance


Super-Res on RGB

For a variety of applications, we want to be able to increase the resolution of images. The ideal algorithm should be fast, and should add sharpness and detail, both at edges and in regions of texture, without adding artifacts. We have made a new version of our Markov network super-resolution algorithm which does not rely on the iterative belief propagation algorithm. The new algorithm has several speed improvements which result in doubling the resolution of an image from 100x100 to 200x200 in less than 2 seconds.

Background and Objectives: For display of images on high resolution display devices, it is desirable to have an algorithm that increases the resolution of the displayed image, so that it has a more pleasing appearance. We would like to achieve higher image quality than our competitors by using a high quality, machine-learning-based image resolution enhancement algorithm.

Technical Discussion: We use a training based approach. We examine many pairs of high resolution, and low resolution versions of the same image data. We divide each image into patches, both high resolution and low resolution patches. We form a training database of 100,000 - 200,000 high and low-resolution patches.

Given a new low-resolution image, we seek to estimate the most probable corresponding high-resolution image. In the training database, there may be several different examples similar to any given input low-resolution patch. However, when we include the boundary conditions corresponding to the neighboring patches that have already been selected by the algorithm, we can find a single best-fitting high-resolution patch for each position in the resolution-enhanced image. We structure the database as a tree, and keep a list of the nearby neighbors to each leaf of the tree. This allows for a fast approximation to the nearest neighbor search that we seek to perform for each patch with the entire dataset. This efficiency results in the dramatic speed up over our previous implementation of the algorithm.

Furthermore, we now convert each input image to a luminance, chrominance color space and only apply super-resolution to the luminance component since this is where the high frequency information mainly resides. This results in a smaller training dictionary by a factor of 1/3, a speed-up by a factor of 3 and fewer artifacts in the resulting super-res image.

Collaboration: We hope to decrease still the time required for processing the images, and to explore the effect of the training set on the image quality of the resolution enhanced image.

Contact: Michael Jones, William Freeman
http://www.merl.com/projects/fastsuperres/

Lab: MERL Cambridge Research
Project Type: Research

Digital Communications

The perspective of today's information society calls for evolution of communication systems and architectural concepts. Although Digital Communications technology has revolutionized telecommunications and computer industries over the last decade, technology development in this field is still rich with innovation. In this landscape, mobile and wireless communications have become the most dynamic sectors of activity.

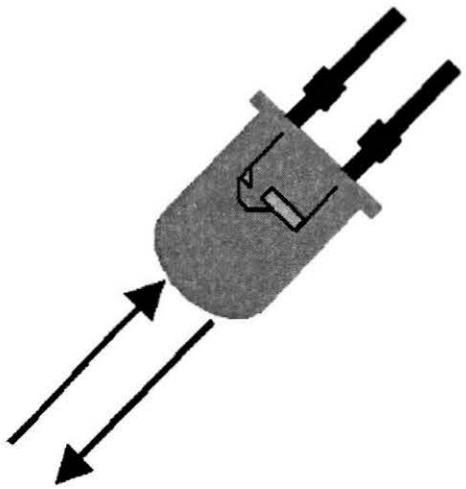
The driving force behind MERL's research depends on the mobility of users. For mobile communications, the question is, "How can the system adapt to the channel conditions so that the data rate and system performance are optimized?" For local area broadband communications, the question is, "What architectures and transmission technologies (optical, cable, wireless, powerline, etc.) are required to increase the bandwidth and services offered to the end user?"

At MERL, we aim to create new concepts and advanced technologies in wireless and wired communications. Participation in development of industry standards is an important part of our work. Our main focus is on three areas:

- Broad Band Mobile Communications. MERL is an active participant in major industry standards for mobile communications protocol, including 3GPP, ITU-T, and ITU-R. We are developing most advanced technologies, such as MIMO/Space-Time Coding, Link Adaptation techniques, for next generation mobile systems.
- Wireless LAN/PAN. MERL is an active participant in wireless LAN/PAN standards, including IEEE 802.11, IEEE 802.15 and 1394 WLAN. We are investigating some key techniques regarding OFDM technology, and developing the new Ultra Wideband radio technology for future high-speed short-range wireless communications.
- 10 Gigabit Ethernet. MERL is an active participant in the IEEE 802.3ae (10 Gigabit Ethernet) standard, and we are developing equalization techniques that will allow new transceiver products to extend distance of the optical link and improve the performance. To transfer the technology, we are also developing the implementation architecture.

Project Descriptions

Sensing and Communication Using Bi-Directional LEDs



Light Emitting Diodes (LEDs) are inexpensive and widely used as light sources. What is less well known is that LEDs are fundamentally photodiodes and as such are also light detectors. We have developed a novel microprocessor interface circuit that can alternately emit and detect light using only an LED, a resistor and two digital I/O pins.

Background and Objectives: This project began as an effort to create a smart backlighting system for television remote controls. A low-power capacitive proximity sensor detects active handling which in turn controls the backlight. To save battery life, the backlight should not be turned on in bright conditions. But adding a separate light sensor would require a new mechanical design for the remote, adding considerable cost. Our solution was to use the backlight LED itself as the light sensor. We developed a simple microprocessor interface technique that uses one additional digital I/O pin, but no other additional components compared to those need to simply light the LED. Since the circuit draws only microwatts of power, it has a minimal impact on battery life.

Technical Discussion: The LED microprocessor interface technique we have developed has far broader implications than simply controlling a backlight. By pointing two LEDs at each other, we can transmit data back and forth. This is trivially easy to do for distances on the order of a centimeter. The result is that almost any LED connected to a microprocessor can be thought of as a potential two-way communications port. We think of this technology as solving "the last centimeter problem" – you have two devices right next to each other, but they have no way to communicate. With LED Comm (LED-based communications), there is a link available, almost for free.

We have been looking into a variety of applications for this very fundamental technology. One opportunity is to turn the power light on appliances into a service port – download status and upload new firmware. Cell phones could swap phones numbers via their backlights. LED-based devices could replace RFID tags in many applications with benefits such as true peer-to-peer communications and a much smaller, far less expensive reader.

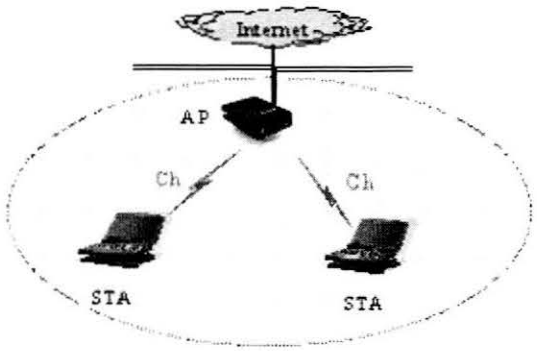
Collaboration: This is a joint effort of MERL Cambridge Systems and Cambridge Research.

Future Direction: Improved data rates, and greater tolerance to ambient lighting

Contact: Paul Dietz, Darren Leigh, William Yerazunis
http://www.merl.com/projects/LEDcomm/

Lab: MERL Cambridge Systems
Project Type: Initial Investigation

IEEE802.11 WLAN Standards Activity



Basic Service Set (BSS)

Wireless LAN (WLAN) is a shared-medium communications network that broadcasts information over wireless links for all stations to receive. IEEE802.11 defines a family of WLAN standards that include original 802.11, 802.11a, 802.11b, 802.11e, 802.11g, 802.11h and 802.11i. In the original 802.11 standard published in 1997, the MAC (Medium access control) and PHY (Physical layer) definitions are described. The other 802.11 standards are either the enhancements to the original MAC for QoS (Quality of service) and security, or the extension to the original PHY for high-speed data transmission. Our activities in IEEE802.11 are focused on contributing to 802.11a and 802.11e, which are the supplement to PHY for high-speed transmission in 5 GHz, and enhancement of MAC for LAN applications with QoS, respectively.

Background and Objectives: WLAN is widely considered to play a major role in wireless multimedia communications that require high-speed transmission with QoS. These technologies need to be standardized. 802.11a defines a high-speed PHY in 5 GHz based on OFDM (Orthogonal Frequency Division Multiplexing) to support up to 54 Mbps data transmission. 802.11e defines an enhancement of MAC to support LAN applications with QoS.

Technical Discussion: 802.11a uses OFDM technology for high data rate transmission. The main idea behind OFDM is to divide the high-speed data stream into several parallel streams of reduced data rate, and to transmit each of them on a separate subcarrier that are made orthogonal to each other. Therefore, spectral overlapping among subcarriers is allowed since the orthogonality will ensure that the receiver can separate the OFDM subcarriers, and a better spectral efficiency can be achieved. In 802.11e standard, a prioritized scheme is used to ensure that high priority users get more bandwidth allocation than low priority users. This is done by a so-called Hybrid Coordination Function (HCF) at MAC. The HCF uses an enhanced contention-based channel access method (EDCF), which operates at stations, concurrently with a polled channel access mechanism operated at AP (access point) to guarantee different data transmissions with required QoS.

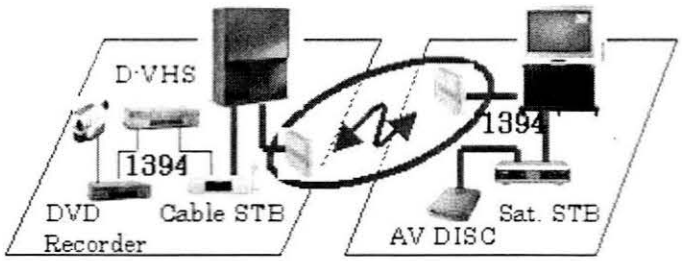
Collaboration: This project is done in collaboration with L-Ji-Se System Development Group.

Future Direction: Contribute to 802.11e standard, and research on algorithms that support multimedia transmission with QoS.

Contact: Daqing Gu, Philip Orlik, Jinyun Zhang
http://www.merl.com/projects/ieee80211b/

Lab: MERL Murray Hill
Project Type: Advanced Development

Ultra Wideband Technologies



Ultra Wideband (UWB) techniques are viewed by many companies as having potential to bring high resolution for tracking and imaging systems and high data rate communication for low cost and low power consumption. A new study group for the High Data Rate of the Wireless Personal Access Network standard (802.15.3a) began to study the possibility to work on a alternative physical layer for Pico nets of 10 meter range and for a minimum data rate of 100Mbps. Because of the impact UWB may have on short-range communications, in this project, we evaluate the capability of this technique to fully understand how to take advantage of it.

Background and Objectives: In February 2002, FCC adopted a First Report and Order that permits the use of UWB devices. Radar and imaging companies were waiting for this decision for several years in order to commercialize their products and to look for new applications, including for short-range communications. The low power consumption of UWB devices opens the door to applications requiring batteries such as storage data devices. Furthermore, the data rates, which are considered by the study group of 802.15.3, allow several new applications like wireless Digital TV, high definition MPEG2 motion picture transferring, DVD playback and DV Camcorder. The main objective is to evaluate possible UWB techniques in order to develop new products for MELCO, which target on short-range wireless communication with high data rate.

Technical Discussion: The first very large category of UWB systems uses train of pulses. The very narrow pulses give an efficient way to combat multipaths in indoor environments and provide a high time resolution for tracking, geolocation and imaging. However the different techniques may be developed under the name Ultra Wideband. Based on the FCC definition, a UWB system uses a very large frequency band (the fractional bandwidth should be greater than 25%). The main difficulty is to generate a signal conforming to the power limitations from FCC and at the same time to be able to achieve high data rate for a reasonable price.

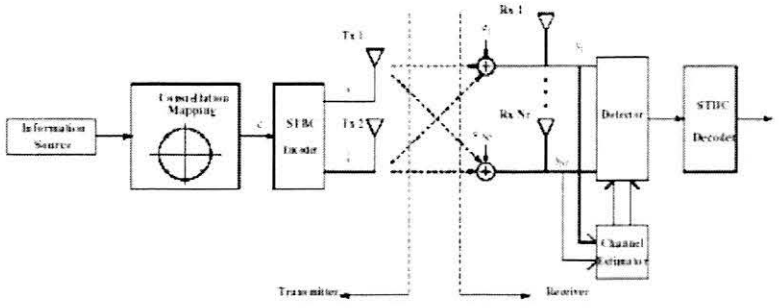
Collaboration: This project is performed in collaboration with the New Jersey Institute of Technologies (NJIT).

Future Direction: The performance analysis of UWB technology will have to be further examined and the transceiver architecture design, pulse generation and antenna design will be investigated.

Contact: Yves-Paul Nakache, Jinyun Zhang
http://www.merl.com/projects/uwb/

Lab: MERL Murray Hill
Project Type: Research

MIMO/Space-Time Coding Systems



The rapid growth in mobile communications leads to an increasing demand for wideband high data rate communications services. Recent research in information theory has shown that large gains in capacity and reliability of communications over wireless channels could be achieved by exploiting the spatial diversity. One appealing technology is called MIMO (multiple input multiple output), which uses multiple antennas at both transmit and receive sides.

Background and Objectives: Space-time block coding (STBC), one of the STC techniques, is used to improve the performance of MIMO systems. It takes advantage of the spatial and temporal diversity as well as coding gain. STBC can provide diversity gain over an uncoded system without sacrificing the bandwidth and increase the effective transmission rate as well as the potential system capacity. The objective of this project is to explore MIMO/STC techniques and develop the key techniques for 3G WCDMA systems and the systems beyond.

Technical Discussion: As a technically and economically feasible method, MIMO antenna processing with interference cancellation is now considered as one of the effective technologies to reach the target of HSDPA in 3G systems. Large capacity in this system is obtained via the potential decorrelation in the MIMO radio channels, where many parallel subchannels are created. Accurate estimation of channel parameters is critical to a reliable recovery of the transmitted signals. Therefore, an effective channel estimation method plays an important role. In this study, the impact of channel estimation errors on the performance of a space-time coded system is investigated when decoding is performed using imperfect estimates of the channel. Combination of turbo coding with the new idea of space-time coding is also explored. In addition, schemes, which adaptively change the coding rate and transmit powers are proven to be able to further boost the performance of STBC systems.

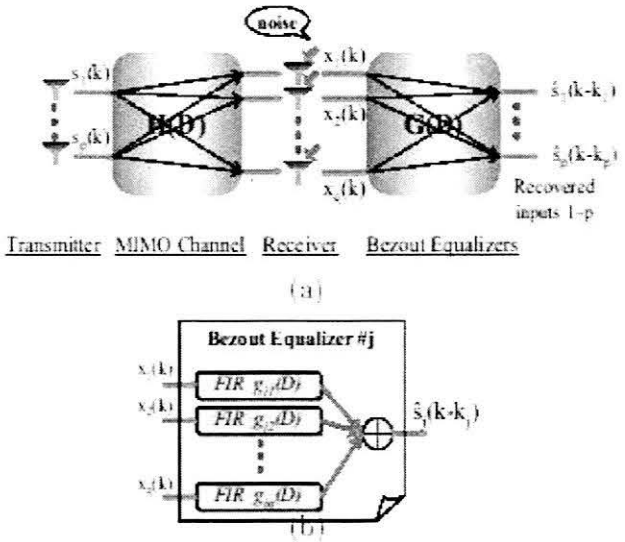
Collaboration: This project is done in collaboration with an intern student from Stevens Institute of Technology in Hoboken, NJ.

Future Direction: MIMO/STC technique is gaining more and more attention. Our focus is to increase the spectral efficiency, improve the system capacity, and provide services with high QoS in a high mobility, multiple access wireless networks.

Contact: Jinyun Zhang
http://www.merl.com/projects/MIMO_STC/

Lab: MERL Murray Hill
Project Type: Research

Bezout Equalizer for MIMO Systems



MIMO (Multiple-Input-Multiple-Output) techniques are viewed by many as having the potential to greatly increase wireless system capacity with multiple antennas in both the transmitter and the receiver end. However, the inevitable ISI (Inter-Symbol-Interference) and ICI (Inter-Channel-Interference) need to be addressed in order to fully explore the advantage of the MIMO systems. We have developed the Bezout equalizer to combat the both interferences through a simple array of linear FIR filters. We air to use this approach to enhance the receiver design and improve the performance.

Background and Objectives: Recently, MIMO techniques attracted lots of attention. One technique, which theoretically achieves channel capacity in MIMO systems, is called BLAST. The original Blast used a cyclic association of data streams, called layers, with transmit antennas, thereby producing an "averaged" channel which is the same for all layers. Difficulties in the realization of the original Blast led to a modified architecture where each layer is associated with a certain transmit antenna. However, in order to achieve the full capacity of the MIMO channel, long data blocks, powerful channel coding, and perfect detection of each layer are required. In addition, in practical systems, the problem of error propagation limits the performance.

Technical Discussion: This work provides a system and method that designs an optimum Bezout space-time equalizer based on estimated MIMO channel characteristics, combines Bezout space-time equalizers with sequential detection and decoding techniques, and processes the input sequences via a layered and pipeline architecture. With a sequential space-time equalizer, an input data stream with a highest signal-to-noise ratio (SNR)is recovered first. Then previously detected transmitting streams are used to reduce interference in subsequent detected input stream. The sequential equalization and detection reduces the number of unknown input streams. The, "excess dimensionality" offered by the increasing asymmetry between the transmitted and received signal space provides the necessary flexibility that improves the capacity of the system.

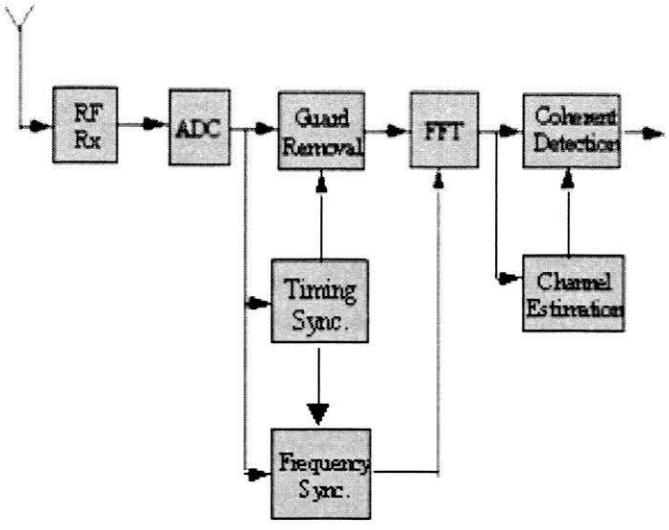
Collaboration: This project is performed in collaboration with Princeton University.

Future Direction: The performance of improvement with Bezout equalizer will be further examined, and detail design together with other receiver components will be investigated

Contact: Jinyun Zhang
http://www.merl.com/projects/bezout_equalizer/

Lab: MERL Murray Hill
Project Type: Research

A Receiver Structure for OFDM-based Wireless LANs



Orthogonal Frequency Division Multiplexing (OFDM) technology has been chosen for high data rate transmissions in many wireless LAN (WLAN) standards such as IEEE 802.11a, HiperLAN/2 due to its spectra efficiency and robustness against multipath and frequency selective fading. In this project, we propose an OFDM receiver structure for OFDM-based WLANs. In the design, a joint timing and frequency synchronization method is used for synchronization. An EM-based (EM: Expectation-Maximization) channel estimator is also used to improve the receiver performance.

Background and Objectives: OFDM technology is highly sensitive to symbol timing and carrier frequency synchronization. Therefore, in the OFDM receiver design, a great attention must be paid to the timing and frequency synchronization. If a coherent demodulation scheme is used, efficient and accurate channel estimation is necessary to compensate for the channel distortion.

Technical Discussion: In our OFDM receiver design, a joint timing and frequency synchronization is used for system synchronization, and EM-based channel estimation is used to compensate for the channel distortion. The timing synchronization consists of two stages: coarse synchronization and fine synchronization. A correlation of the training symbols is used for coarse synchronization. Then, up-sampling and digital interpolation filtering are employed to provide additional resolution to estimate the correct timing more accurately. The frequency offset estimator uses MLE (Maximum Likelihood Estimation) algorithm to estimate the frequency offset. The EM-based channel estimation algorithm consists of two iterative steps: the expectation step (E-step) and the maximization step (M-step). The proposed OFDM receiver structure improves the system performance.

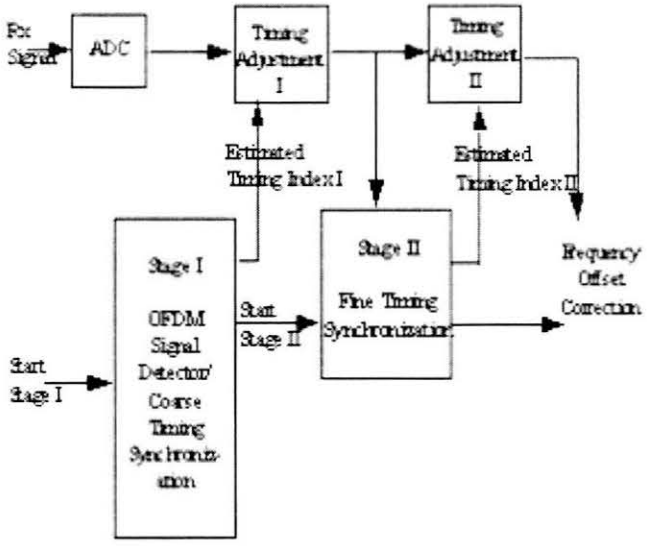
Collaboration: This project is done partially in collaboration with L-Ji-Se System Development Group and Dr. Makoto Miyake in Johosoken.

Future Direction: We continue to refine our proposed OFDM receiver structure, and seek its application to L-ji-se 802.11a chipset development.

Contact: Daqing Gu, Jinyun Zhang
http://www.merl.com/projects/ofdm_receiver/

Lab: MERL Murray Hill
Project Type: Research

Timing Synchronization for OFDM-Based WLAN



Due to its spectra efficiency and robustness against fading, OFDM (Orthogonal Frequency Division Multiplexing) technology has been widely adopted as a modulation scheme for many WLAN (Wireless LAN) standards such as IEEE802.11a and HIPERLAN2. The performance of an OFDM system heavily relies on its timing and frequency synchronization. In this project, we propose a new timing synchronization technique for OFDM-based WLANs. The proposed method can detect an OFDM symbol, and accurately estimate the beginning of the OFDM symbol simultaneously with low-complexity.

Background and Objectives: OFDM is highly sensitive to synchronization errors: timing error and frequency error. In order for an OFDM system to work, the receiver and the transmitter have to be synchronized. This includes timing and frequency synchronization. Timing synchronization involves in finding the exact timing instant of the beginning of each OFDM symbol. Unless the correct timing is known, the receiver cannot remove the cyclic prefixes at the right timing instant of the symbol and correctly separate individual symbols before computing the FFT of their samples.

Technical Discussion: Our proposed timing synchronization consists of two stages: coarse synchronization and fine synchronization. In the coarse synchronization, a received preamble of WLAN data frame is passed through a correlator. Due to the identical time-domain symmetry of the short training symbols in the preample, the normalized correlated output is larger at the correct timing than at an incorrect timing. Thus, the correlated output forms a peak at the correct timing. In the fine synchronization process, the up-sampling and digital interpolation filtering are used to provide additional high-resolution samples for the training symbols. Thus, the resolution of the correlation is increased. In this way, the timing of an OFDM symbol can be accurately estimated.

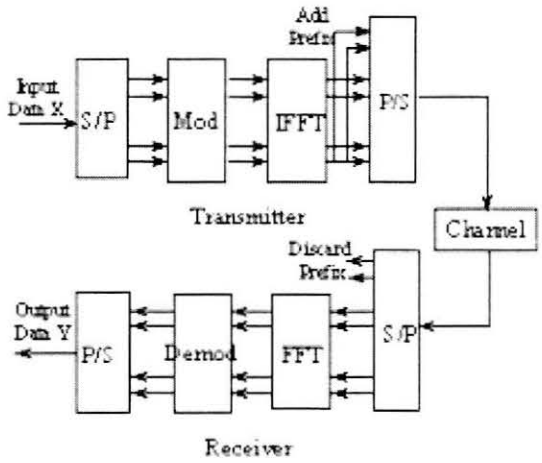
Collaboration: This project is supported by Melco CR&D

Future Direction: This project is complete. The next step is to seek its application.

Contact: Daqing Gu, Jinyun Zhang
http://www.merl.com/projects/ofdm_timing/

Lab: MERL Murray Hill
Project Type: Research

EM-Based Channel Estimation and Signal Detection for OFDM Systems



Channel estimation and signal detection plays an important role in the communication receiver design if a coherent demodulation is used. In general, it is not possible to make an accurate data decision at the receiving end unless a good channel estimate is available. In this project, we propose an EM-based channel estimation and signal detection algorithm for OFDM (Orthogonal Frequency Division Multiplexing) systems. The maximum likelihood estimate of CIR (Channel Impulse Response) is obtained by using channel statistics via EM (Expectation-Maximization) algorithm. The proposed method significantly reduces the BER (Bit Error Rate) and improves the system performance.

Background and Objectives: In OFDM link, data are modulated on the subcarriers by some form of PSK or QAM. In general, each received symbols show random phase shift and amplitude change. To cope with these, an appropriate signal detection method is needed. One of the signal detection methods is coherent detection in which the estimates of the reference phases and amplitudes are used to determine the transmitted data. EM-based channel estimation algorithm provides an efficient and accurate method to estimate these references.

Technical Discussion: The EM algorithm is an iteration method for finding maximum likelihood estimates of system parameters in a broad range of problems. In each iteration, EM algorithm consists of two steps: the expectation step and the maximization step. The expectation step is performed with respect to unknown underlying parameters, using the current estimate of the parameters, conditioned upon the incomplete observations. The maximization step then provides a new estimate of the parameters that maximizes the expectation of log likelihood function defined over complete data, conditioned on the most recent observation and the last estimate. These two steps are iterated until the estimated values converge.

Collaboration: This project is done in cooperation with Princeton University, and supported by Melco CR&D.

Future Direction: We continue to research on OFDM technology, and place our focus on PAPR (Peak-Average-Power-Ratio) reduction topics.

Contact: Daqing Gu, Jinyun Zhang
http://www.merl.com/projects/OFDM/

Lab: MERL Murray Hill
Project Type: Advanced Development

HSDPA in W-CDMA

HSDPA (high speed downlink packet access) is a key feature in the 3rd generation wireless communication standard W-CDMA (wideband CDMA (code division multiple access)), which was launched in late 2001 by NTT DoCoMo in Japan. This technology provides high data rate transmission (up to 8-10 Mbps, and 20 Mbps for MIMO systems) in CDMA downlink to support multimedia services. We researched and developed the key technologies on fast link adaptation for HSDPA, such as AMC (adaptive modulation & coding), interference cancellation, scheduler design, and MIMO (multiple input multiple output), to enhance MELCO's IPRs and prepare the essential technologies for 4th generation wireless communication.

Background and Objectives: The objective of this project is to develop the key technologies on fast link adaptation for HSDPA and to enhance MELCO's IPR portfolio on 3G systems. In 3GPP (3rd generation partnership project) standards, Release 4 specifications provide efficient IP support enabling provision of services through an all-IP core network and Release 5 specifications focus on HSDPA to provide data rates up to approximately 10 Mbps to support packet-based multimedia services. MIMO systems are the work item in Release 6 specifications, which will support even higher data transmission rates up to 20 Mbps.

Technical Discussion: HSDPA is a packet-based data service in W-CDMA downlink with data transmission up to approximately 10 Mbps over a 5MHz bandwidth. This system is evolved from and backward compatible with Release 99 WCDMA systems. To reach this target, some key technologies are used for HSDPA: AMC, HARQ (hybrid automatic request), MIMO, fast cell search, and advanced receiver design. In general, all HSDPA users share the channel in both time and code domains. Adaptive modulation and coding is used to support multiple rate transmission for different types of multimedia services. Due to the high data rate transmission, the trade-off between complexity and capability of the UE (user equipment) and base-station becomes an important issue. The research in MERL is focused on the MIMO systems, advanced receiver design and scheduler algorithm development to reach fast link adaptation for HSDPA.

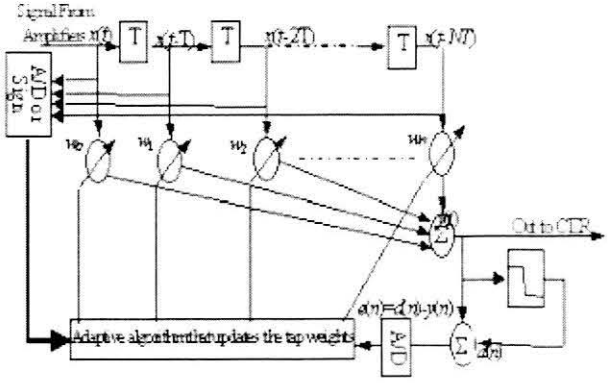
Collaboration: MHL collaborated with Johosoken for the development of HSDPA systems.

Future Direction: MIMO system enhancements for HSDPA

Contact: Jyhchau (Henry) Horng
http://www.merl.com/projects/HSDPA/

Lab: MERL Murray Hill
Project Type: Advanced Development

IEEE 802.3ae Equalizer Project



MELCO is currently developing a 10Gigabit per second optical communications transceiver module. This transceiver is to be based on the new IEEE 802.3ae Ethernet standard that is to be complete later this year.

In this project, MERL-MHL has been investigating the use of equalization technology after optical to electrical conversion. Equalizers can be used to overcome dispersion introduced by the optical fiber and other system components and improve system performance as well as increase transmission distance.

Background and Objectives: Future data networks are evolving towards a more IP based architecture, in addition the coverage area/transmission distance of these networks are also expanding. The IEEE 802.3 body is developing new standards to increase both the speed and reach of its "Ethernet" standard. The market for such devices is expected to grow and equalization is a key technology that will allow MELCO to differentiate its products from other 10Gigabit Ethernet suppliers.

Technical Discussion: The new Ethernet standard defines both multimode fiber and single mode fiber for use as physical media. In each case the fiber type determines the dispersion mechanisms, intermodal dispersion in multimode fibers and chromatic and polarization mode dispersion in single mode fibers. The dispersion mechanisms result in inter-symbol interference (ISI) and increased bit error rate at the receiver, eventually ISI limits the distance that signals can propagate and still be recovered.

While adaptive equalizers have been used in many communication systems in the past, their use in optical data communications is still relatively new. The challenge here is the high speed of operation. Even with today's advances in VLSI and high-speed circuitry, the implementation of advanced digital equalizers at 10Gbps is very difficult. To eliminate the need for digital implementation we have proposed a hybrid analog and digital approach for the equalizer's implementation. In addition simplified weight updating techniques can further reduce the implementation complexity. Simulation results have shown that this adaptive equalizer architecture is capable of compensating for the ISI expected from both multimode and single mode fibers. Currently, it is believed that an equalized version of 10 Gigabit Ethernet can achieve increases in the maximum transmission distance.

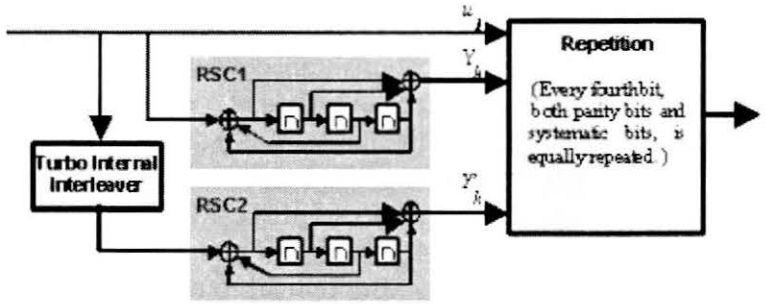
Collaboration: MERL-MHL is actively collaborating with L-ji-se System Development Group and MEUS.

Future Direction: Future work will focus on integration of proposed equalizers with MELCO transceiver design and improving equalizer performance.

Contact: Philip Orlik
http://www.merl.com/projects/10GbEqualizer/

Lab: MERL Murray Hill
Project Type: Advanced Development

Turbo Codes for HSDPA



Structure of Turbo Encoder using repetition method

The release 99 version of the 3GPP standard uses a rate 1/3 turbo encoder. A new rate ¼ turbo code based on the repetition of bits from the 1/3 code has been proposed for use in the HSDPA (High Speed Down-Link Packet Access) enhancement to be published early in 2002. In this project MERL-MHL has been investigating the performance within an HSDPA system of rate ¼ turbo codes based on rate 1/3 codes and bit repetition. In addition, we have investigated techniques to improve BER performance of sub-optimal decoding algorithms like the SOVA as well as VLSI implementation issues.

Background and Objectives: There are two reasons behind the use of bit repetition for the HSDPA enhancements. First is it reliance on an already standardized 1/3 rate turbo code. Secondly, is the ability to use a less complex rate 1/3 turbo decoder in the receiver/handset. However, the bit repetition method will offer weaker BER performance when compared to more traditional rate ¼ turbo codes based on puncturing from lower rate codes.

Technical Discussion: It is well known that the MAP (Maximum A Posterior) decoding algorithm is the optimal decoding algorithm for turbo codes. However, it has a drawback in its complexity and cost of implementation. Sub-optimal algorithms such as the SOVA (Soft Output Viterbi Algorithm) have been developed to simplify decoder implementation. Also it has been shown that simple modifications to these sub-optimal decoding algorithms can improve BER performance. Specifically, scaling the extrinsic information that is feedback to the decoder can reduce the BER of the decoding algorithm.

We have found that the use of a simple scaling of the extrinsic information in the decoding process can improve the BER performance of SOVA based turbo decoders for the HSDPA ¼ rate turbo code. In an HSPDA system were the preference is to use the repetition method, the scaling technique is shown to help minimize the penalty incurred by using this sub-optimal encoding scheme. Additional investigations on optimal methods to combine the repeated bits were carried out.

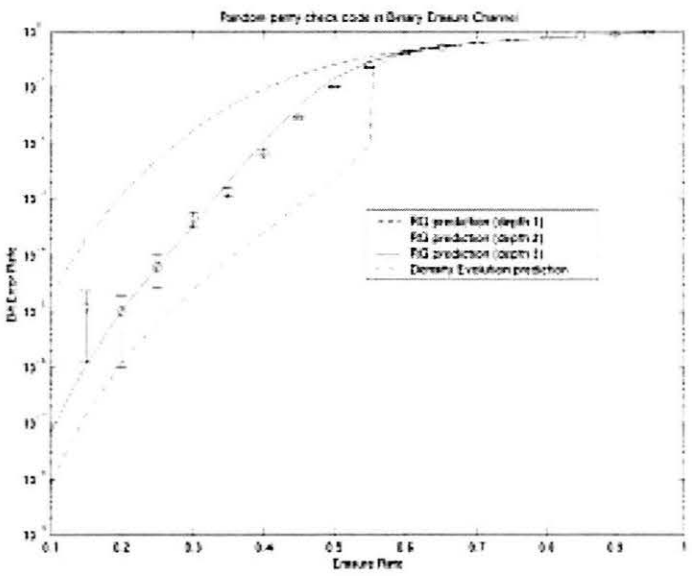
Collaboration: MERL-MHL is actively collaborating with Dr. Duanyi Wang, Dr. Hisashi Kobayashi of Princeton University and Dr. Lei Cao of University of Missouri-Columbia, and Dr. Makoto Miyake in Johosoken.

Future Direction: This project is completed.

Contact: Philip Orlik
http://www.merl.com/projects/turbo_codes/

Lab: MERL Murray Hill
Project Type: Research

Analyzing and Designing Error-Correcting Codes



A fundamental problem in information theory is the analysis and design of optimal and efficiently decodable error-correcting codes of a given block-length and rate. The discovery of turbo codes in 1993 and the re-discovery of Gallager codes soon thereafter stimulated considerable research into iterative decoding of codes on graphs. One can now argue that in the infinite block-length limit, we understand how to approach the Shannon limit. For practical applications, however, one must work with codes of finite block-length, for which existing theory is far from sufficient. We have developed new techniques to analyze and design finite-blocklength codes which can be decoded using iterative message-passing algorithms.

Background and Objectives: The best-known technique to analyze the performance of iterative belief propagation decoding is the "density evolution" method. However, this method presumes that the "messages" passed in the decoding algorithm are statistically independent, and is therefore only valid for the infinite block-length limit of certain codes such as irregular Gallager codes and their relatives. Our objective is to devise an efficient technique for computing the entire performance curve for an arbitrary parity check code of finite block-length. This performance curve can then be used to estimate the performance of a code at very low error-rates where simulation is impossible, or as an objective function in a search over codes.

Technical Discussion: We have so far developed two new techniques. The first is based on the "renormalization-group" approach from physics: the idea is to continually replace an error-correcting code with a simpler error-correcting code that has nearly identical performance, until the code is reduced to a small enough size that its performance can be computed exactly. The second technique, which we call "projective algebra," is a generalization of density evolution that exactly accounts for the statistical dependencies between messages in the same loop.

Collaboration: with Erik Sudderth (MIT) and Jean-Philippe Bouchaud (CEA-Saclay)

Future Direction: Our work, and that of others who have considered iterative decoding of finite-blocklength codes on the binary-erasure channel, shows that the best way to predict the performance of a code is to measure the size of its minimal "stopping sets" - those collections of bits which will stop decoding if they are erased. We are currently developing techniques to efficiently search for finite-blocklength codes with large minimal stopping sets.

Contact: Jonathan Yedidia
http://www.merl.com/projects/ECC/

Lab: MERL Cambridge Research
Project Type: Research

Graphics

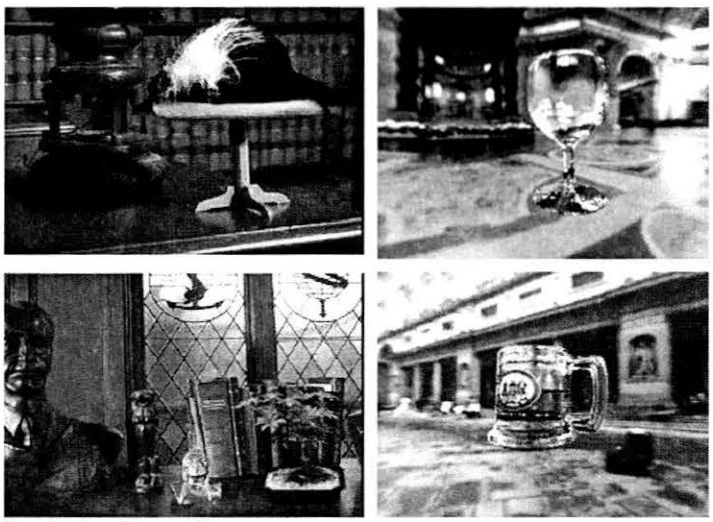
Computer Graphics includes the computer-based creation, storage, manipulation, and display of geometric models and 3D images. Thanks to tremendous advances in graphics hardware and software, Computer Graphics has pervaded our everyday life, becoming an enabling technology in applications ranging from art and entertainment to science and engineering.

MERL has a strong history in Computer Graphics research, with participation in the premiere industry conferences and with diverse research projects. Our focus during the past year can be broadly classified into 3 main areas:

- Model acquisition. We have several projects that use digital cameras for model acquisition. These include a custom, image-based scanning system that acquires both shape and appearance using multiple cameras and computer-controlled lights, a method for acquiring a 3D scene via user-assisted segmentation of multiple photographs taken from different viewpoints, and a method for acquiring detailed geometric texture from pairs of photographs taken under different lighting conditions. Additional research includes a technique to improve model reconstruction from range scanners.
- Modeling. We are investigating new digital representations for shape including Adaptively Sampled Distance Fields (ADFs), point-based representations, opacity hulls, and reflectance fields. We are also researching systems and interaction techniques for model creation. These include a sketch-based system for creating smooth, free-form 3D shapes from 2D gestures, and a high-end digital sculpting system for creating detailed models for CAD/CAM and entertainment.
- Display. Our research includes methods for displaying Computer Graphics, ranging from algorithms for high quality rendering to projector systems. Research in projectors includes a self-correcting projector that pre-warps the projected image to ensure that it is rectangular and aligned with the display surface and a system that provides apparent 3D motion by projecting real-time graphical animation on static 3D objects.

Project Descriptions

3D Images



See Color Figure 8

We built a system for automatically acquiring high quality graphical models of objects that are extremely difficult to scan with traditional 3D scanners. The acquired models can be placed in arbitrary new environments. This system is the first to acquire and render fuzzy, transparent, and transparent 3D objects, such as feathers or a glass of beer, from arbitrary viewpoints under novel illumination.

Background and Objectives: Reproducing real objects as believable 3D models that can be placed into arbitrary synthetic environments has been a longstanding goal in computer graphics. Traditional 3D scanning techniques, although capable of accurately capturing 3D shape, are not well suited to capture complicated object appearances of, for example, highly specular, fuzzy, or transparent objects. We have developed an image-based 3D scanning system that is very robust and capable of fully capturing 3D objects that are difficult if not impossible to scan with existing scanners. We also have developed new rendering techniques to display the acquired objects with high image quality.

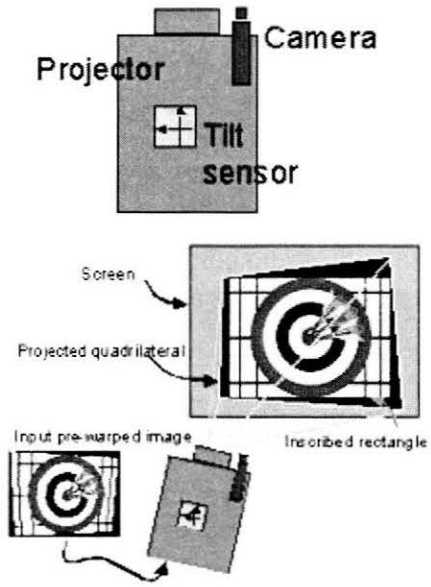
Technical Discussion: The system consists of turntables, a set of cameras and lights, and monitors to project colored backdrops. We use multi background matting techniques to acquire alpha mattes and images of the object from multiple viewpoints. The alpha mattes are used to construct the opacity hull, a new shape representation. The opacity hull is defined as the visual hull of the object with view-dependent opacity. It enables visualization of complex object silhouettes and seamless blending of objects into new environments. Our system also supports relighting of objects with arbitrary appearance using surface reflectance fields, a purely image-based appearance representation. Our system is the first to acquire and render surface reflectance fields under varying illumination from arbitrary viewpoints. We have built three generations of digitizers with increasing sophistication.

Collaboration: This project is a collaboration between MERL, MIT, and the Swiss Federal Institute of Technology (ETH) in Zürich, Switzerland.

Future Direction: We are working on compression and progressive transmission of 3D images while improving the rendering speed.

Contact: Hanspeter Pfister, Paul Beardsley, Remo Ziegler **Lab:** MERL Cambridge Research
http://www.merl.com/projects/3Dimages/ **Project Type:** Research

Smart Projector



See Color Figure 9

We have developed a smart projector that can automatically sense its orientation with respect to a vertical or horizontal surface. The goal is a self-correcting projector to generate a rectangular image of known aspect ratio, even when aimed at an arbitrarily inclined planar surface. Furthermore, for a vertical or near-vertical display surface, the sides of the rectangle should be aligned appropriately with the world vertical. The projector is augmented with a rigidly attached camera and a tilt sensor. We exploit the homography between the projected image and the displayed image for keystone correction. By sensing the relative orientation between the screen and the projector, we can pre-warp the image so that it appears rectangular when projected. In addition to efficient pre-warping rendering methods, we have developed new techniques to calibrate the camera, projector and tilt-sensor system.

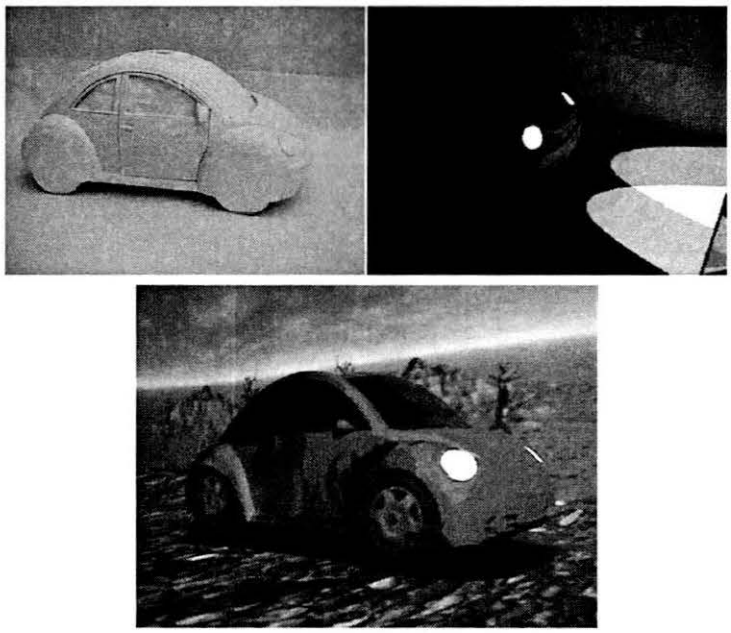
Background and Objectives: A self-correcting projector that always displays a rectangular image is a desirable feature of every intelligent projector. We can eliminate the cumbersome procedure of aligning a projector with respect to a screen and make it very easy to install and use. Our approach is unique due to two features. First, we can project on blank walls i.e. without fixed markers to create two-dimensional Euclidean frame of reference. Second, we use a single self-contained device. Hence, we avoid the use of devices that are external to the projector, such as calibrated cameras mounted at known locations.

Technical Discussion: The tilt-sensors detect the out-of-vertical-plane orientation of the projector lens. Further, the camera detects the shift in the horizontal plane. The measured orientations lead to the well known 8-parameter planar homography which can be used represent the relationship between original image and the keystoned displayed image. We use an inverse of that relationship to pre-warp the image. Currently, the pre-warping is achieved in real time using 3D graphics hardware.

Collaboration: We are investigating techniques to improve the interface to such intelligent devices.

Contact: Ramesh Raskar, Paul Beardsley, Jeroen van Baar **Lab:** MERL Cambridge Research
http://www.merl.com/projects/SmartProjector/ **Project Type:** Initial Investigation

Apparent Motion with Shader Lamps



See Color Figure 10

We have developed a set of techniques to create apparent motion on real objects by projecting real-time animations. This project is based on an extension of the Shader Lamps concept. By introducing a set of techniques and applications that involve projectors, non-planar surfaces and computer graphics animations, our goal is to demonstrate the use of projectors in a new type of display medium.

Background and Objectives: Animations in movies delight the audience with moving characters but they remain on a flat 2D screen. Physical dioramas, on the other hand, are detailed, three-dimensional and allow physical interaction but they are static. Our techniques to combine the two in some limited cases. We illuminate static physical models with projectors. The images are generated with real time three dimensional computer graphics. Our system demonstrates various visual effects such as non-photorealistic shading, apparent motion and virtual lighting on a toy-car model.

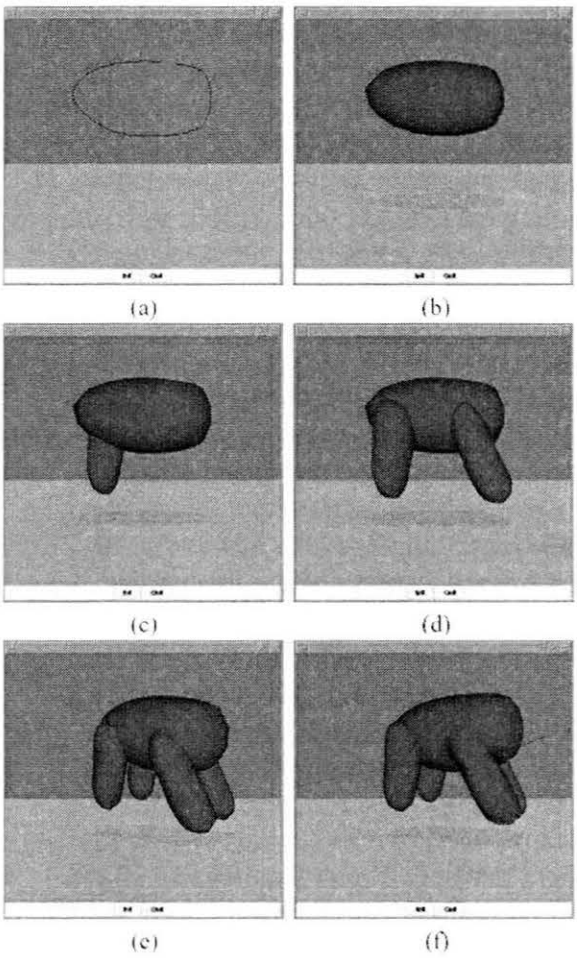
Technical Discussion: Projector can be used to change the surface appearance of neutral colored physical models by illuminating them with rendered images. The projectors are treated as 'Shader Lamps' [Raskar et al. 2001]. In this type of projector-based augmentation, complete illumination of complex 3D shapes is made possible by rendering geometric models of those shapes from multiple overlapping projectors. The geometric models are pre-authored with detailed color, texture and material properties to reproduce desired surface appearance. Although, reproduction of surface reflectance is well studied, simulating apparent motion from static models remains unexplored.

A great deal of perceptual and physiological research has been conducted to discover the properties of the mechanisms for motion analysis by humans. The illusions are created due to the need to sense retinal motion, and analyze it as quickly as possible. We combine such illusions with 3D dimensional animations to create apparent motion effects that would be otherwise difficult to achieve.

Contact: Ramesh Raskar, Remo Ziegler
http://www.merl.com/projects/MotionSL/

Lab: MERL Cambridge Research
Project Type: Research

Sketching 3D Shapes with Variational Implicit Surfaces



See Color Figure 11

We have developed a system that allows the creation of certain freeform objects with a sketching interface. "Sketching" means gestural marks in 2D to model shapes. We have observed that a particular representation for underlying geometry - the variational implicit surfaces - allow a unified method for a collection of free-from modeling operations. In this work, we have created operations for easy editing of shapes and developed techniques based on variational implicit surfaces. With the ability to convert the models to and from the traditional polygonal format, the underlying representation remains transparent to the user. The techniques are ideal for creating smooth free-from shapes.

Background and Objectives: With the advent of sketch-based methods for shape construction, there's a new degree of power available in the rapid creation of approximate shapes. Just as in children's sketching books, one is taught to first draw the general forms of things starting from simple pieces (cylinders, spheres, cones, blobs etc.), and then to draw a more careful outline and erase the underlying shapes, we provide the user the opportunity to roughly sketch out shapes and then modify them to provide a final form.

Technical Discussion: We provide three basic operations: (i) "inflate" 2D stroke into a 3D shape (ii) "merge" two or more 3D objects into a single object with a guidance stroke and (iii) "modify" by oversketching an existing shape.

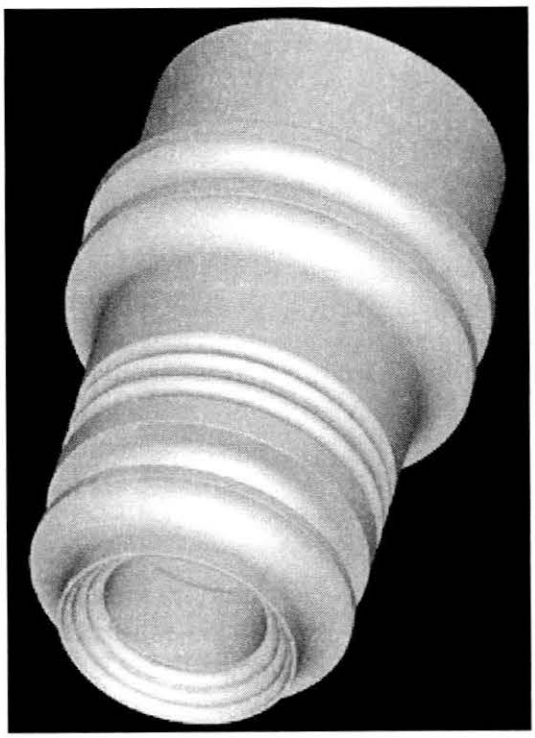
The basic idea is to simply remove or add constraints on an implicit surface to create, merge or modify the shapes. The constraints are inferred from user-supplied strokes. In addition, we use an automatic hierarchy generation method to represent the parent-child relationship between objects as they are created.

Collaboration: Brown University

Contact: Ramesh Raskar
http://www.merl.com/projects/sketching/

Lab: MERL Cambridge Research
Project Type: Research

Adaptively Sampled Distance Fields (ADFs)



ADFs are a new digital representation of shape with several advantages over existing approaches. They provide efficient and accurate representation of both smoothly curved surfaces and surfaces with fine detail. ADFs have the potential to impact many diverse industries including CAD/CAM (simulation, path planning, and verification for milling precision parts), Entertainment (building models for games and movies), Fonts (high-quality display of letterforms for PDAs), Visualization (volumetric visualization of molecular structure), 3D Scanning (3D models from image or range data), 3D Printing (rapid prototyping), and Color Management (projectors, PDAs, monitors, and printers).

Background and Objectives: Our objectives include fundamental research, incorporation of this research into a product-worthy C library ready for commercialization, development of a comprehensive patent portfolio, and collaboration with key industrial players to refine and expand the vision for ADFs.

Technical Discussion: A distance field is a scalar field that specifies a distance to a shape, where the distance may be signed to distinguish between the inside and outside of the shape. ADFs consist of adaptively sampled distance values, organized in a spatial data structure, with a method for reconstructing the distance field from the sampled distance values. This approach permits the accurate and compact representation of fine detail and smooth surfaces, together with efficient processing. ADFs allow: the representation of more than the surface (interiors and exteriors); the compact representation of sharp features and organic shapes; smooth surface reconstruction; trivial inside/outside and proximity testing; fast and simple CSG operations; fast geometric queries such as closest point; and efficient computation of surface offsetting, blending and filleting, collision detection, morphing, and rough cutting.

Collaboration: PTC (ProEngineer CAD/CAM software), Think3 (PTC competitor), Industrial Light and Magic (movies: Star Wars, Jurassic Park, ...), WetaFX (Lord of the Rings movie trilogy), NOVA documentary series, and MELCO's Factory Automation group.

Future Direction: Fundamental research and technology transfer

Contact: Sarah Frisken, Ron Perry
http://www.merl.com/projects/adfs/

Lab: MERL Cambridge Research
Project Type: Research

Net Services

Net Services technology leverages the ubiquity and accessability of information provided by the Internet. Internet-ready devices are showing up in automobiles, hospitals, the mobile workplace, and nearly every room in the home. Soon the personal computer will not be the primary means for people to access the Internet. This makes Internet Software a rich area of research and development, as the Internet continues to expand into diverse arenas such as public, private, B2B, B2C, wireless, entertainment, and location aware information services.

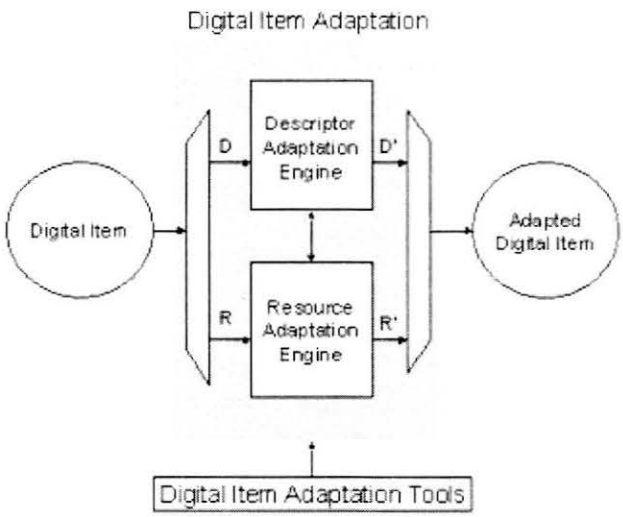
MERL is focused on developing middleware and applications that enable new types of Internet-based services and software products. Our most significant efforts are in the areas on Mobile Agents and Wireless Internet Networking:

- Mobile Agents: MERL as developed a mobile agent technology, called Concordia (page 117), which is currently licensed by several customers. It is a framework for developing and managing network-efficient mobile agent applications useful for accessing information anytime, anywhere and on any device supporting Java. The Concordia system allows custom mobile agents to travel across the Internet, where they can interface with backend applications, databases, or other mobile agents. We are continuing to develop the Concordia framework, making it useful for a wider range of applications, including Enterprise Application Integration.
- Wireless Internet Networking. MERL is developing PAMLink (Personal Access Mobile Link, formally known as TWINet), a wireless Internet networking framework that allows for easy, ubiquitous access to personalized information services and content using a personal wireless communication device. PAMLink, a technology applied to several MERL projects, is a cell-phone-centric networking technology that turns common consumer appliances into an Internet appliances.

PAMLink is an evolving project and in FY2002, MERL has added a new significant feature that has been dubbed PAMLink-21. PAMLink-21 is a combination of the PAMLink architecture, which is used as a network platform, with the emerging MPEG-21 standard. The MPEG-21 standard provides the "glue logic" for transporting content such as MPEG video streams to the user. MPEG-21 describes entities in the system such as the user terminal's capabilities, the network capabilities, the server capabilities, and the choice of available content itself.

Project Descriptions

MPEG-21 Standards Activity



One of the goals for MPEG-21 (MPEG: Moving Pictures Experts Group) is to enable transparent access to multimedia content. At this time, there are seven different parts of the standard defined by the current workplan. The fundamental unit of transaction in MPEG-21 is called a Digital Item. Simply stated, Digital Items contain a rich set of content with associated descriptions, and they are configurable to various conditions. Our activities in MPEG-21 are focused on contributing to the development of Part 7, Digital Item Adaptation. The goal of this part is to standardize tools to assist with the adaptation of Digital Items. This part is schedule to become an International Standard in 2003.

Background and Objectives: Today, many elements exist to build an infrastructure for the delivery and consumption of multimedia content. There is, however, no "big picture" to describe how these elements, either in existence or under development, relate to each other. The vision for MPEG-21 is to define a multimedia framework to enable transparent and augmented use of multimedia resources across a wide range of networks and devices.

Technical Discussion: To enable transparent access to content, it is essential to have available not only the description of the content but also a description of its format and of the usage environment in order that content adaptation may be performed to provide the User the best content experience for the content requested with the conditions available. While the content description problem has been addressed by MPEG-7, the description of content format and usage environments has not been addressed and it is now the target of MPEG-21 Digital Item Adaptation. Among the tools targeted for standardization include usage environments descriptions, which include descriptions of the terminal, network, user and natural environment.

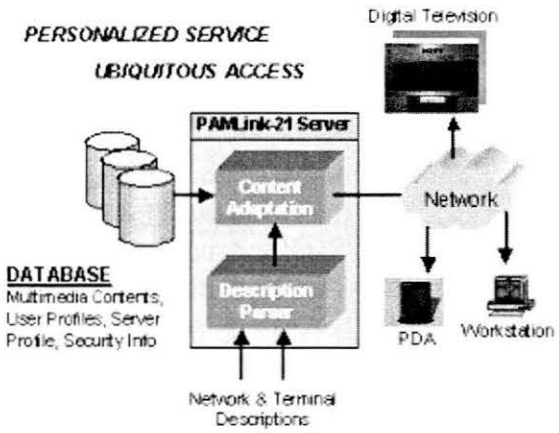
Collaboration: This project is done in collaboration with the Multimedia Information Coding & Transmission Technology Department at Johosoken

Future Direction: Contribute to the development of MPEG-21, especially Part 7: Digital Item Adaptation. Also, implement key software components to realize a compliant MPEG-21 system.

Contact: Anthony Vetro
http://www.merl.com/projects/mpeg21/

Lab: MERL Murray Hill
Project Type: Research

PAMLink-21



The primary aim of the PAMLink (Personal Access over Mobile Link) project formally known as TWINet is to develop an end-to-end connectivity solution for Internet Appliances and Internet services. With this in mind, PAMLink services are characterized by a highly personalized experience for the individual user, where the services and information are aggregated to meet individual needs at any given moment. This is very much in-line with the vision of the emerging MPEG-21 (Moving Pictures Experts Group) standard, which is to enable transparent access to multimedia content. PAMLink-21 merges these two initiatives together with the goal of developing key components for an MPEG-21 system and integrating them with a rich multimedia-enabled infrastructure.

Background and Objectives: This project concentrates on merging technology that has been developed to gain personal access from mobile devices with the emerging MPEG-21 standard.

Technical Discussion: The current thrust of this project is to use the PAMLink infrastructure and the MPEG-21 standard to stream multimedia content to various user's terminals each having different capabilities. Various descriptors, which are in the process of being standardized by MPEG-21, are used to describe the producer, the consumer, the network delivery system, as well as the expressions of various rights a user may have to the content. The server, whose responsibility is to manage the services and deliver content, is comprised of a content adaptation engine and various parsers to interpret the MPEG-21 descriptions. The client has an integrated HTML/XML browser and MPEG-4 Player, which has been implemented on a WinCE platform. At this current stage, the content delivery network is based on RTP/UDP/IP (RTP: Real-time Transport Protocol; UDP: User Datagram Protocol; IP: Internet Protocol).

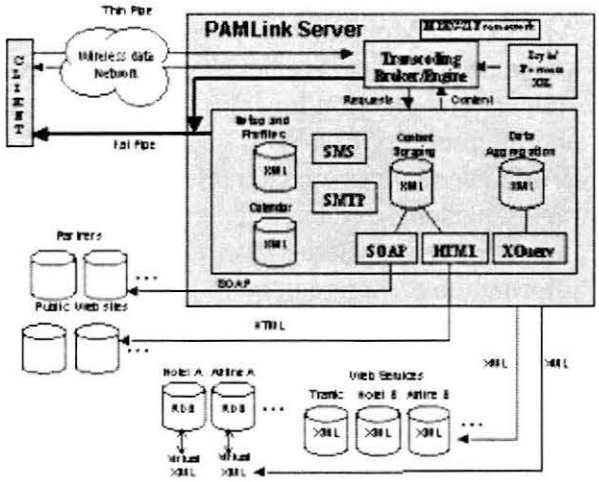
Collaboration: Certain parts of this project are done in collaboration with the Multimedia Information Coding & Transmission Technology Department at Johosoken, such as the MPEG players, video delivery mechanisms, and MPEG-21 components.

Future Direction: Enhancing the operation of the server is one primary goal in the near future. This will include adding mechanisms for QoS provisioning and traffic flow management. Also, we plan to enhance the content delivery network to also include transmission over IEEE 1394.

Contact: Johnas Cukier, Anthony Vetro
http://www.merl.com/projects/PAMLink21/

Lab: MERL Murray Hill
Project Type: Research

PAMLink Database and Web Services



PAMLink (old name: TWINet) is a wireless networking framework that allows for easy, ubiquitous access to personalized information services and content using a personal wireless communication device, using as client any compatible appliance such as a TV set, PDA or a car navigation system for the rendering. The users access their personalized services such as calendar, bank account information, picture album, etc. at home on a TV set, or in the car through the sound system. The services are configured for each user and for each client device. The server side stores personal profiles and services, manages content brokerage, transmission and storage, and interacts with the clients.

Background and Objectives: The PAMlink Server constitutes the server side of the PAMLink framework. It manages the user and device profiles, information gathering and storage for the services, and the transmission to the clients. The objective is to provide to the users high value services that they can access anywhere on any compatible appliance.

Technical Discussion: The PAMLink server side comprises of databases that store user and services information, applications that provide services, and interfaces for the administration of the user profiles and the overall system. The applications gather content (e.g. traffic, status of corporate voicemail, etc.) data via the Internet based on user requirements. The server interacts with a wide variety of sources accessed via diverse protocols ranging from existing HTML, SOAP, and SMTP, to XML-based Web Services. Currently supported services include Intranet content such as calendar, email, and corporate voice mail status in addition to Internet content like weather, traffic, stocks, lottery, to-do list, photo gallery. Available services and device types are stored in the database. The users' database store personal information, preferences for each selected device and for each selected service. Currently, we target the TV and the car navigation system as the primary client devices.

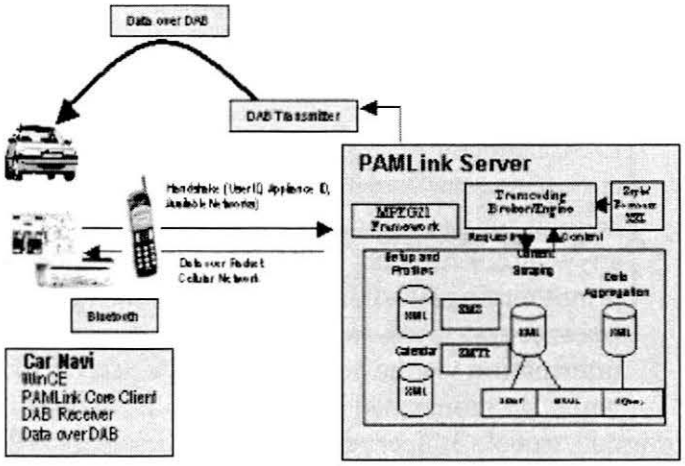
Collaboration: Johosoken, Sentansoken, MEAA (Mitsubishi Electric Automotive America), MDEA (Mitsubishi Digital Electronics America)

Future Direction: Various significant improvements are underway. (1) Integration with Web Services technologies based on XML (2) Adoption of standards such as MPEG-21 (currently under development) for description of terminals, capabilities, and user environments. (3) Transmission of content via alternative paths generally supporting higher speeds. (4) Scalability of the server.

Contact: Fernando Matsubara, Johnas Cukier
http://www.merl.com/projects/TWINet-server/

Lab: MERL Murray Hill
Project Type: Advanced Development

Wireless Connections to Car Navigation Systems



With the explosive growth of the Internet for both business and personal needs, easy-to-use connectivity solutions are required. PAMLink formally known as TWINet can be used to gain access from various locations without the need to involve specialized equipment such as a PC, complicated wiring, or limitations of a small display in current mobile devices. PAMLink easily turns any appliance such as TVs and car telematics systems into Internet connected devices by using a cellular phone.

Background and Objectives: This project concentrates on the car telematics/navigation systems used with PAMLink on a Windows CE platform.

Technical Discussion: The key feature of Wireless Connections to Car Navigation is the ability for the user to bring his personal network environment to the car. Here the user is quickly brought to a familiar environment that spans across the office, home, and home away from home environment. The user also has access to pertinent data relative to the automobile experience. In other words, the PAMLink system is sensitive to the user's environment as well. The system identifies four basic entities: the user, the appliance (e.g. car telematics, TV, etc.), the available communications channels, and the services available to the user. For the automobile, the typical components available in the car might be a: car navigation system, global positioning system, and various car sensors (engine diagnostics, speed, tire sensors, fuel level indicator, etc.). The availability of other components that are not typical today would include local storage for services such as multimedia files (movies, games, maps, guidebooks, etc.).

The automobile interface is another issue. The main concern for the driver is that he is not distracted while using the system. For that reason, voice interaction may play a key role for the automobile interface. User interface components include: text-to-speech, voice command recognition and a simple one to three button interface that is easily accessible by the driver. The passenger interface may include a richer set of components such as: high-resolution screens/touchscreens, keyboards, mice, game controllers, infrared remote controller, etc.

Collaboration: Sentan-Soken(Sanken), MEAA (Mitsubishi Electric Automotive America)

Future Direction: In the future, wireless wide-area (Digital Audio Broadcasting) and narrow-area (IEEE 802.11 standard) broadcasting will be available to the motoring public. The automobile system should take advantage of both developments for streaming multimedia to the car.

Contact: Johnas Cukier
http://www.merl.com/projects/TWINet-car-nav/

Lab: MERL Murray Hill
Project Type: Advanced Development

PAMLink Digital Television

The mobile access to the personalized information and Internet is becoming a requirement. There is a need to provide a mobile network solution that delivers quality services. PAMLink (formerly called TWINet) DTV is an easy-to-use answer. The only device needed is a cellular phone, which is in hand any time, anywhere. PAMLink DTV supports intelligent programming functionality. User can navigate Internet and access to the personalized information via the high quality DTV screen. No wires, no more small display of mobile device, no extra skills needed to set up.

Background and Objectives: PAMLink DTV is an application of the PAMLink architecture. The objective is to provide a mobile access to user's personalized information and Internet by simply utilizing the cellular phone and DTV. The interconnection between cellular phone and DTV is as easy as plugging the cellular phone into a DTV docking station. Once cellular phone plugged into docking station, DTV is connected to PAMLink server via the wireless channel. The PAMLink server provides user with the personalized information and Internet service. The cellular phone is used as an identifier for user's personal data. When un-plugged, the cellular phone and the DTV can be operated independently. All PAMLink data will be removed from DTV immediately.

Technical Discussion: To provide a personalized mobile access solution, utilize the high quality DTV display and overcome the small display size of cellular phone, we have proposed and implemented PAMLink DTV to leverage MELCO strengths in DTV. The serial technology is used to connect DTV and the cellular phone. The wireless upstream channel is used for sending IP data to service provider. For downstream data, the wireless downstream channel or other available fast channel may be adapted. The user's personal information is associated with the cellular phone and therefore no explicit dialogue is required for the access. User can edit his/her profile to add or remove services anytime. With PAMLink DTV, TV program and PAMLink data can be displayed simultaneously. PAMLink data can be displayed with transparent background or opaque background. The different display font and color can also be chosen. User is notified automatically if there is a new message. User may view PAMLink data by just pressing DTV remote control buttons.

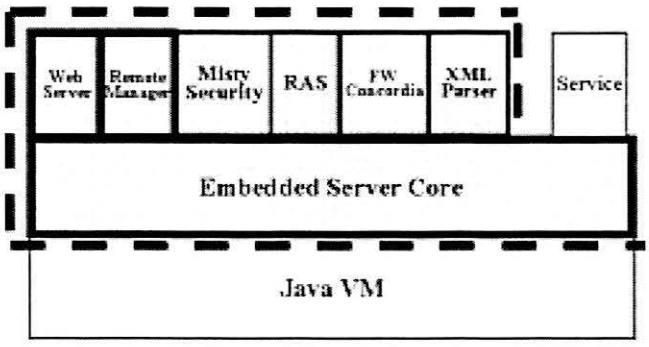
Collaboration: Mitsubishi Electric Digital Equipment America, Inc.

Future Direction: We will also explore the use of Bluetooth and eventually wireless 1394 for the future generations of PAMLink DTV.

Contact: Jianlin Guo
http://www.merl.com/projects/TWINet-DTV/

Lab: MERL Murray Hill
Project Type: Advanced Development

Concordia for Power Systems Business Units



A major success story for the Concordia mobile agent program has been its adoption by MELCO's power systems business units as a core technology in their solutions for the electric power industry. For the past few years, much of the Concordia development has focused on meeting their requirements and on effectively transferring this technology to MELCO for formal product customization, maintenance, and support in Japan.

Background and Objectives: The Concordia mobile agent technology has been well received by MELCO's power systems business units and has become a core technology in their systems solutions for the electric power industry. In the past few years, the Concordia program has focused much of its development efforts specifically to meet their requirements. Furthermore, we have conducted thorough technology transfer to these business units in order that formal product customization, maintenance, and support for Concordia can be established in Japan. We continue to support this productization effort with periodic product enhancements.

Technical Discussion: Some of the important Concordia product features that have been developed specifically for the power systems business units include: 1) streamlining of the Concordia core, 2) enhancing Concordia for wide area network support, and 3) enhancing Concordia security features.

The effort on streamlining the Concordia core and reducing its memory footprint resulted in the subsequent development of Featherweight Concordia for embedded devices. The work required to make Concordia more suitable to wide area networks and the Internet included improving the error handling capabilities of the Concordia runtime and making it much easier for enterprise users to transmit Concordia agents through existing corporate firewalls. Work is ongoing for both efforts as customer needs are identified.

Collaboration: The Concordia program has been working very closely with MELCO's power systems business units for the past three years. MERL has also been working closely with the Texas A&M University and PSERC (Power Systems Engineering Research Center) to explore new applications of agent technology in the power systems domain.

Future Direction: We will continue to serve in the role of being the primary mobile agent technology provider to MELCO's business units.

Contact: David Wong, Frederick J. Igo, Jr., Derek Schwenke **Lab:** MERL Cambridge Systems
http://www.merl.com/projects/concordiaSanden/ **Project Type:** Advanced Development

Social Net

Social Net is a novel interest-matching application that uses patterns of collocation over time to infer shared interests between users. It demonstrates new possibilities and methods for using the capabilities of mobile devices equipped with RF communications.

Background and Objectives: Wireless, peer-to-peer RF-communication (e.g., Bluetooth) is an increasingly accessible option when designing mobile devices and applications. A recent trend in both the commercial and research worlds has been to outfit pocket-sized devices with RF-based communications so that applications can continually broadcast the users presence, while simultaneously detecting others nearby. We consider such applications opportunistic, mobile applications, because they opportunistically perform a service or function when detecting others in the physical world. Social Net explores new strategies for using mobile devices equipped with RF-based communications.

Technical Discussion: We propose new strategies for using mobile devices equipped with RF-based communications. These devices can be configured to automatically detect when people with similar devices are collocated, yielding data that can offer rich contextual information when analyzed over time. We believe that analyzing patterns of collocation, frequency of collocation, time of day of collocation, and duration of collocation, all yield insights into the type and quality of relationships held between the user and other people encountered. We have designed and implemented an application of these strategies in Social Net, a novel interest-matching program that infers shared interests between people who are frequently collocated, over time. When Social Net infers two strangers share interests, it seeks out a mutual friend of the pair, and if one is found, she receives a message suggesting she introduce the pair.

A key aspect of social net is keeping a "human in the loop" to address social concerns and issues of scalability. Circumventing existing social practices can be problematic. Therefore, rather than attempting to directly mediate an introduction, Social Net relies on a person's social network to locate an individual to perform the introduction of the two strangers. This requirement introduces trust and accountability into the process, while assigning the most critical social roles to the person, not the technology.

Collaboration: MERL-CS and MERL-CR are working in collaboration with Georgia Tech on this project.

Future Direction: A pilot and small-scale field study have been conducted. Additional experiments are needed to evaluate the efficacy of this approach. In addition, Social Net is one sample application; we would like to develop other context-aware applications that explore these new strategies.

Contact: Kathy Ryall, Darren Leigh
http://www.merl.com/projects/snet/

Lab: MERL Cambridge Systems
Project Type: Research

Digi-Koi: A Game for Cell Phones

See Color Figure 12

We have developed a game for cell phones with small color screens that is a simulation based on the popular Japanese pastime of breeding koi. Our digi-koi look like real koi and our simulated genetics exhibit many of the characteristics of real koi genetics. Unlike more conventional computer games, the digi-koi game does not have an intrinsic goal. Although we leave open the possibility of contests to breed digi-koi with certain characteristics, we think that the primary motivation for playing is likely to be aesthetic pleasure: a typical player will probably just want to breed beautiful digi-koi. A koi trading mechanism allows players to augment their gene pool by acquiring digi-koi with genetic characteristics that are absent from their pond. The trading mechanism is very simple; however, the trading process can be complex in a social sense and adds significant richness to the game.

Background and Objectives: Cell phones offer game designers unique challenges and opportunities. A typical cell phone has limited memory capacity and restricted input capabilities. However, recent cell phones have added (relatively) strong output capabilities: a 132x162-pixel screen with 12-bit color; and Short Messaging Service (SMS). We do not yet have an adequate repertoire of interface widgets for performing typical interactions on cell phones and other small screen devices. What are the best ways to support browsing, selection, text entry, etc., on a device with a small screen and no dedicated pointing or scrolling device? We do not yet know what types of games will be the most popular on cell phones. Will ports of existing PC or PDA games dominate or will a new class of compelling games emerge that are tailored to cell phones? Through this project, we contribute to the ongoing debates on both of these questions.

Technical Discussion: The distinctive features of our game include two accommodations to limitations of cell phones. Firstly, the genetic makeup and lineage of a digi-koi is represented in fewer than 70 bits, thereby minimizing data storage and facilitating the electronic transfer of koi. Secondly, we have developed new interface widgets, tailored to the constraints of a cell-phone, for handling the browsing and selection operations needed for the game.

Collaboration: We plan to generalize digi-koi's interface widgets for small screen devices and compare them to existing interaction techniques. Additionally, we are exploring licensing the game with a variety of game companies.

Contact: Joe Marks
http://www.merl.com/projects/digikoi/

Lab: MERL Cambridge Research
Project Type: Initial Investigation

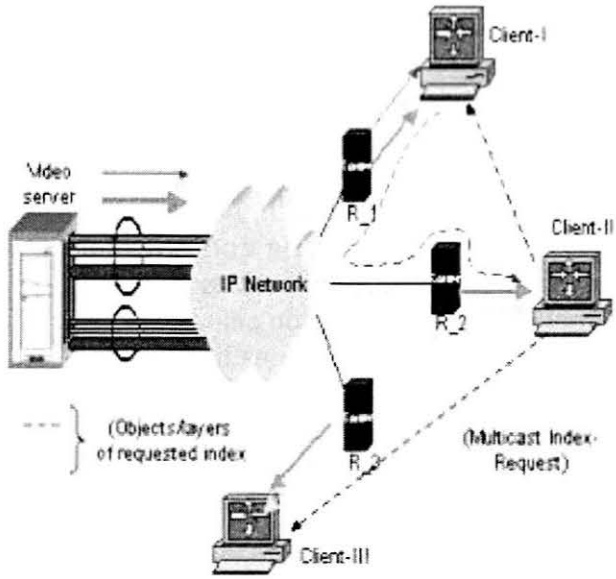
Networks

Computer networks have already become a pervasive substrate in our daily life from at-work, on-the-go, to the home front. As digital text and voice data communication are already a common practice, consumers' apetite has started to demand visual streaming multimedia data. Efficient delivery of streaming media over the Internet and wireless networks present many new research challenges.

Today, more often than not, the end-to-end network communication path from source to destination include a multiplicity of different network technologies ranging from wired or wireless Ethernet LAN, the Internet, to celluar networks. The current Internet and wireless networks only offers best effort service, thus does not guarantee or provide quality of services (QoS) for multimedia applications, many of which embody timing constrains in its value. Therefore, temporally changing network conditions (e.g., bandwidth variations, network induced delay, etc.) can adversely impact the quality of the content to be delivered. A large body of MERL's network research addresses the various multimedia transmission, packetization and transportation issues for next generation beyond 3G wireless networks and the Internet.

Project Descriptions

QoS Traffic Management



Efficient delivery of streaming media over the Internet presents many challenges. The current Internet only provides best effort service, and it does not guarantee or provision quality of services (QoS) (Quality of Service) for multimedia applications. Therefore, temporally changing network conditions (e.g., bandwidth variations, network induced delay etc.) may adversely impact quality of content to deliver. On the other hand, the second complication is due to that the features, which describe the variations in content of the multimedia, are generally imprecise. This research aims to develop techniques in managing network resources to increase statistical multiplexing gain in multicasting of FGS (Fine Granular Scalable) video with high quality delivery. These techniques include effective router bandwidth and buffer management, content adaptation to available resources, and proxy server design etc.

Background and Objectives: The IETF Integrated Services model includes additional components beyond those used in the best-effort model such as packet classifiers, packet schedulers, and admission control. Routers, switches, hubs are the entities where resource allocation decisions have to be made, and the decisions have to be enforced in a network. The goal is to provision application level QoS, to achieve high utilization and statistical multiplexing gain especially in video streaming and multicasting.

Technical Discussion: We classify traffic management studies into two as media-unaware and media-aware. Media-unaware approaches assume no a-priori traffic knowledge, unlike the latter. First, we have developed a method for media-unaware dynamic resource allocation for a VBR stream. The method exploits multiple predictors and uses a finite state decision mechanism in which state-to-state transitions occur with interrupts generated according to delay and available bandwidth measurements. Less queue size occupancy and higher utilization are achieved after less number of reallocations than those of other methods in the literature.

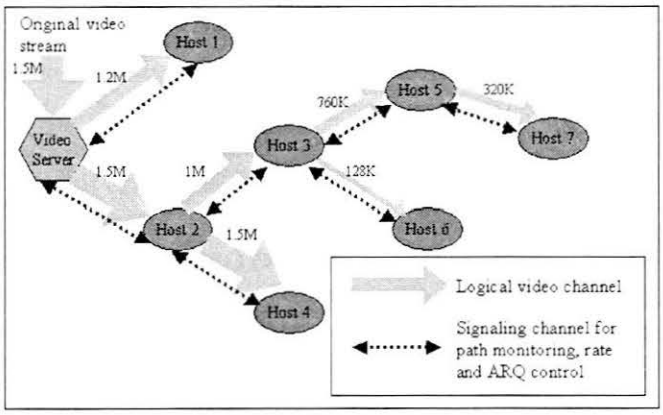
Collaboration: A part of this project will be collaborated with NJIT.

Future Direction: Development of media aware resource management techniques (e.g., proxy caching, layer add/drop etc.) for multicasting of FGS video, and content management in video surveillance to alleviate network load.

Contact: Zafer Sahinoglu
http://www.merl.com/projects/qostrafficmanagement/

Lab: MERL Murray Hill
Project Type: Initial Investigation

Adaptive Video Over End-System Multicast



End-System Multicast is regarded as a promising overlay network architecture for multimedia group communication over the Internet. This project investigates how to integrate adaptation capability from the application level and flexible multicast functionality from End-System Multicast for pre-stored video streaming applications. Our approach can support video streaming to heterogeneous user devices with different computing and bandwidth capabilities.

Background and Objectives: Multicast is considered as an effective communication support for multi-party applications since it can greatly save network bandwidth. However, traditional IP Multicast is plagued with concerns pertaining to scalability, network management, deployment and support for higher layer functionality such as error, flow and congestion control. A new overlay multicast architecture called End-System Multicast was proposed recently. Pushing multicast functionality to edge or host of the network, End-System Multicast brings more flexibility to add higher layer functionalities for multimedia communications. Rate adaptation in multimedia is regarded as a necessary mechanism to handle network and users' heterogeneity, and fluctuations of available network bandwidth. This project proposes a new approach of adaptive video multicasting to heterogeneous users.

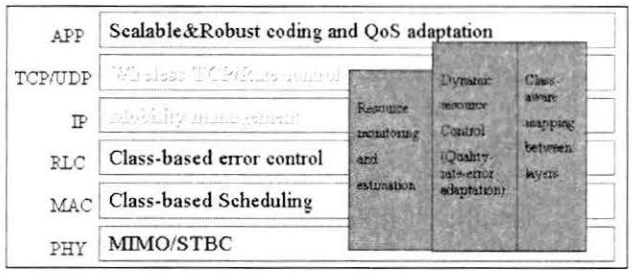
Technical Discussion: In this project we investigate how to achieve rate adaptation for pre-stored video streaming in the architecture of End-System Multicast. The key idea is to take advantage of both application adaptation and overlay multicast adaptation to obtain optimal end-to-end video transmission performance. A fast and simple packet-level rate adaptation mechanism was proposed to overcome the drawbacks of the layered video multicast approach. With this packet-level mechanism, only one network session is required for multiple layers of the same video and fine rate adaptation at intermediate nodes can be done easily combined with End-System Multicast. We also study how to extend the previous work of End-System Multicast to support application adaptation well.

Collaboration: Because End-System Multicast is more appropriate for small group applications, our proposed approach also aims to applications with video streaming to small group receivers. Our future work includes how to support large group receivers in applications such as Internet TV to hundreds thousands of users.

Contact: Huairong Shao, Chia Shen
http://www.merl.com/projects/videomcast/

Lab: MERL Cambridge Research
Project Type: Research

Adaptive QoS Control for Wireless Multimedia



This project explores how to support dynamic resource adaption and system optimization for multimedia applications in the next generation wireless network. We propose dynamic multi-class quality-rate-error adaptation spanning all layers to support multimedia services. Inter-layer resource interaction and coordination is also studied in this project.

Background and Objectives: It can be foreseen that multimedia services and applications such as video cellular phone and web browsing will proliferate in the next generation wireless network. Quality of Service (QoS) provision for multimedia services is regarded as a challenging task due to special characteristics of wireless links such as fading and mobility. Adaptive resource control and management is a promising approach to wireless multimedia. This project aims to provide effective resource adaptation techniques for multimedia over wireless networks.

Technical Discussion: Our adaptive QoS control solution spans from application layer down to physical layer and includes both network (relay stations) and end-systems (handheld devices). The key idea is dynamic multi-class quality-rate-error adaptation. At the application layer, we focus on scalable and robust coding schemes and mapping of QoS parameters. Rate control and mobility support are considered in the network and transport layer. We propose new scheduling mechanism and admission control scheme for the MAC layer. Hybrid ARQ and tradeoff between diversity and multiplexing gain for multiple transmit and receive antennas (MIMO channel) are also considered in the physical layer.

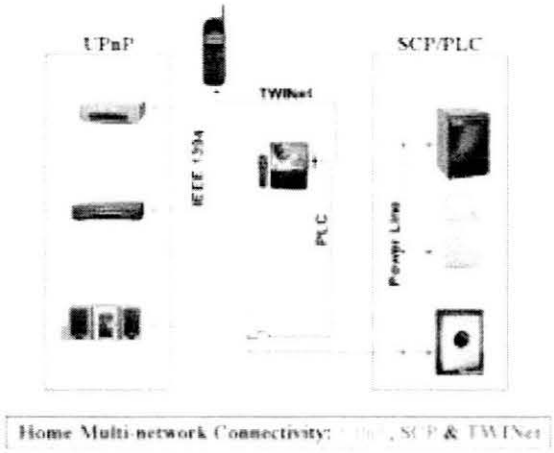
Collaboration: J. Zhang, D. Gu and P. Orlik from MERL-MHL on scheduling mechanism.

Future Direction: This research is currently being carried out.

Contact: Huairong Shao, Chia Shen
http://www.merl.com/projects/adaptiveQoS/

Lab: MERL Cambridge Research
Project Type: Research

Universal Plug and Play (UPnP)



Home Multi-network Connectivity: [illegible], SCP & TWINet

Universal Plug and Play (UPnP) is fast emerging as a widely accepted standard for connecting home devices via TCP/IP connectivity. UPnP technology aims at providing the common person, who is not a computer-networking expert, with a simple, reliable, and an efficient solution to most, if not all, networking and connectivity aspects in a home environment. UPnP offers a solution to many networking aspects including, but not restricted to, managing connections (a UPnP enabled device joins a network automatically once it is plugged in), managing data contents (content directories), controlling device operations, and data streaming and rendering.

Background and Objectives: UPnP seems very promising among many emerging home networking technologies. It aims at providing connectivity in homes to a common non-technical user. Development of other technologies such as Simple Control Protocol (SCP), Power Line Carrier (PLC), Havi, Jini, and IEEE 1394 implies that future home networks will be hybrid systems. Microsoft is especially supporting the two complementary technologies, that is, the UPnP and SCP. Mitsubishi is very well positioned in making strides in these areas and get benefited by the opportunities offered by these developments.

The objective of this project is to develop prototype UPnP and SCP/PLC systems along with the bridging technology.

Technical Discussion: A UPnP network typically consists of a control point (CP) and UPnP-enabled devices. The CP provides a user interface for controlling the operations of the network. The UPnP stack consists of a several well-tested protocols, such as Simple Service Description Protocol (SSDP), Simple Object Access Protocol (SOAP), and General Event Notification Architecture (GENA), HTTP and TCP/IP.

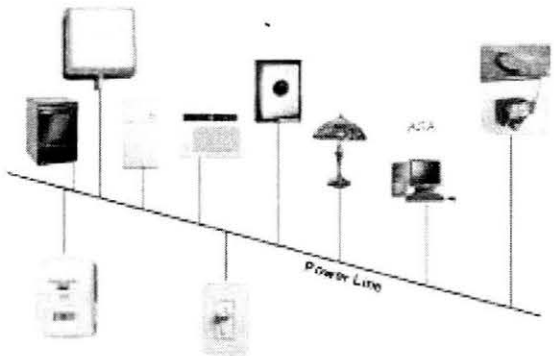
MHL is has implemented the UPnP stack and developed a demo consisting of several UPnP devices, a pair of PLC devices, and a bridge. The demo was designed to show the Mitsubishi DTV working as the center of home entertainment as well as home networking. The demo was placed in CEATEC 2001 show.

Collaboration: MHL is worked very closely with MELCO Han-pon (L-Ji-Se) on MCPLC part of the demo.

Future Direction: Home networking offers a big opportunity for manufacturers of UPnP devices. The future Mitsubishi DTVs should be UPnP-enabled. More than that the DTV should act as the center of home entertainment and a controlling station across hybrid home networks.

Contact: Fernando Matsubara, Ghulam Bhatti, Johnas Cukier **Lab:** MERL Murray Hill
http://www.merl.com/projects/upnp/ **Project Type:** Advanced Development

Simple Control Protocol (SCP)



A Typical SCP/PLC Home Network

Simple Control Protocol (SCP) is a home networking protocol, proposed by Microsoft Corporation, which allows the manufacturers to produce small and intelligent devices. Initially, it will use Power Line Carrier (PLC) as its physical layer. In an apartment building complex, all devices connected to the power line make a single physical network. But the devices in each apartment make a logical network. The system provides mechanism for defining logical networks and provides security. But the offered bandwidth (7.5 kbps) is a very limited. An SCP network is shown in the diagram above.

Background and Objectives: The home networking is an important emerging area with vast opportunities for growth and revenue. The SCP is a complementary technology to Universal Plug-n-Play (UPnP) networking. The UPnP uses the IP as its basis. The SCP targets the class of devices with limited computing power, low bandwidth, and relatively low price. This class of devices mostly consists of home appliances, sensors and light bulbs. The main objective of networking of these devices is to provide automated operational control, diagnostic, and remote updating of firmware, etc. The project aims at integrating SCP/PLC technologies with Mitsubishi M16C processor. That chip can be supplied to device manufacturers at a very low cost.

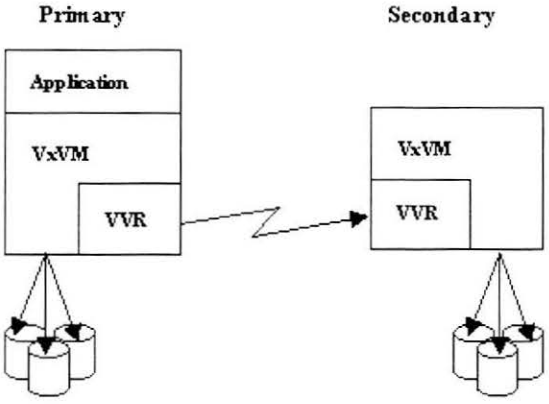
Technical Discussion: The current implementation of the SCP/PLC runs on an ARM7 based platform and is still under development. MHL is porting the code to an M16C based platform. The significant architectural differences of the two processors mean that the bulk of the porting effort will consist of implementing the lower level processor specific code. The current implementation takes really good advantage of several processing modes with separate register banks available on ARM7. The M16C processor, however, has only two register banks. MHL is working hard and confident of having the SCP running successfully on the new platform as soon as possible.

Collaboration: MHL is working closely with MELCO Han-pon (L-ji-se) on the SCP project.

Future Direction: The SCP is one of many emerging home networking technologies. It must be able to interoperate with other technologies such as UPnP, HAVi, etc. Developing the bridging technologies will be very important to make SCP accepted in the market.

Contact: Fernando Matsubara, Ghulam Bhatti, Johnas Cukier **Lab:** MERL Murray Hill
http://www.merl.com/projects/SCP/ **Project Type:** Advanced Development

Network Replication



The Network Replication project is a cooperative effort with Veritas Software to provide network replication capability to their popular Volume Manager product, VxVM. Our technology has been licensed by Veritas for their Veritas Volume Replicator product (VVR).

The replication engine of VVR offers an ideal solution for the remote archiving of storage such as databases or file systems through its ability to replicate a virtually unlimited number of related data volumes while maintaining consistency. Replication uses standard networks without proprietary hardware and is highly resistant to system and network failure.

Veritas Software is a market leader in storage management software.

Background and Objectives: The Network Replication project evolved from an R&D exploration of network distributed storage. Establishment of our relationship with Veritas Software refined this objective to network replication of Veritas' logical volumes.

Technical Discussion: Veritas' customers use VxVM to protect their data from media failure by creating local mirrors, but remote mirrors were needed to protect against system and infrastructure failure. VVR replicates groups of logical volumes, maintaining consistency among them during replication. Changes to any member of the volume group are transparently captured and replicated to one or more remote locations.

VVR has flexible replication and configuration characteristics. It can replicate synchronously or asynchronously. Feedback allows input flow rates to be throttled if necessary to match available network bandwidth. It can simultaneously replicate volume groups to remote sites while acting as a receptacle for volume groups replicated from other sites and it can support an unlimited number of volume groups. It can replicate a volume group to multiple destinations, each with independent replication characteristics and latencies. Recent work has focused on providing support for TCP and otherwise improving network performance.

Collaboration: Since 1997, CSL has worked exclusively with Veritas Software in the development of VVR, deriving market direction and market access from this relationship.

Future Direction: The next year's work is likely to focus on ease-of-use issues, including configuration, administration, and performance monitoring. Work on performance improvements in some areas is also planned.

Contact: David Rudolph
http://www.merl.com/projects/netrep/

Lab: MERL Cambridge Systems
Project Type: Advanced Development

Spoken Language Interfaces

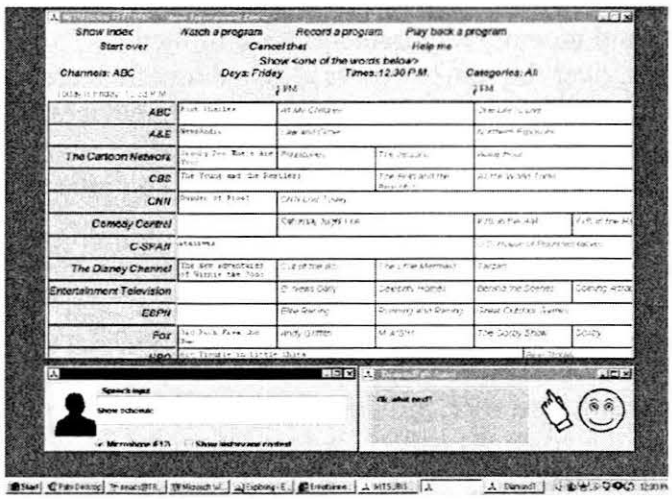
Speech recognition and speech synthesis technologies have been available in crude form for nearly two decades. However, their processing and memory requirements have limited applications to a few niche markets. Meanwhile, over the same two decades, a dramatic increase in the complexity of consumer electronics, cellular telephones and mobile computing devices has created a market need for Spoken Language Interfaces in order to simplify the user interface and free-up the hands and eyes. So while speech recognition and synthesis technologies continue to make incremental improvements, the potential for a revolution in applications of Spoken Language Interfaces looks promising.

At MERL, our approach to research in Spoken Language Interfaces takes three directions:

- Speech-Centric Devices. Most applications of speech technology take the approach of layering a speech interface on top of existing hardware, providing the user the option of pressing buttons with either the finger or the voice. At MERL, we are interested in creating new devices designed with Spoken Language Interfaces in mind. We believe that the result will be improvements in industrial design, hardware costs, and interface accuracy.
- Conversational Speech Interfaces. At MERL, we are applying principles from human collaborative discourse theory to build a conversational framework on top of which we can layer speech interfaces. With this framework, we are able to build into the software system some understanding of the task at hand. We believe that this will make conversational, natural Spoke Language Interfaces more accurate, robust, and useful for real products.
- SpokenQuery. Information Retrieval (e.g. Inktomi, Google, AllTheWeb) has become an important and very lucrative technology. However, the current set of search engines all require typed input, and there are many situations where a keyboard is not acceptable. MERL's goal is to enable Information Retrieval using only spoken queries. Instead of typing the query, the SpokenQuery user verbally describes the desired information. The ability to access information in the automobile, on the cellphone and on PDAs combined with an interface that is hands and eyes free will enable large new markets.

Project Descriptions

User Study of a Collaborative Spoken-Language Interface



We have studied the effects of requiring a user to speak to a collaborative agent using a small, constrained subset of English. Our study used a prototype spoken-language collaborative agent for operating a personal video recorder(PVR), implemented using the COLLAGEN middleware. The user communicated his goals to the software agent in simple spoken English, and the agent asked for more information when necessary.

Background and Objectives: Speech recognition technology has matured dramatically in the past few years, with the first generation of products using embedded speech recognition now coming to market. These products typically support only a very small set of commands. Our focus is a second generation of spoken-language interfaces, which will be more collaborative and conversational. One approach to maintaining low speech error rates in such systems is to use a small, constrained subset of a natural language.

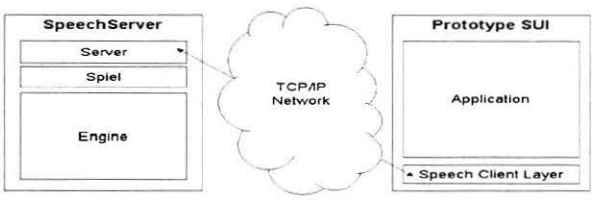
Technical Discussion: Our study compared the task success that users had when the constrained subset of English they must use is either always available ("persistent help") or available only when requested ("non-persistent help"). All users were able to complete their tasks in a reasonable amount of time. Persistent-help users had about an 80% success rate in their first session using the system, while non-persistent help users had a 60% success rate. In subsequent sessions, where persistent help was not available, the non-persistent-help users' success rate improved to eventually reach 80%, while the persistent-help users' success remains unchanged. It appears that persistent-help users did not retain knowledge of the constrained subset language as readily as did the non-persistent-help users.

Collaboration: This study will inform our design of future spoken-language collaborative interfaces. In particular, we plan to continue to use constrained subset languages, because (1) users can succeed with their tasks using such interfaces, and (2) the use of such interfaces reduces recognition error rates.

Contact: Candace Sidner
http://www.merl.com/projects/subset/

Lab: MERL Cambridge Research
Project Type: Research

SpeechServer



The SpeechServer is a plug-in software component which provides speech recognition capabilities for rapid prototyping of new speech user interfaces.

Background and Objectives: It has historically been very costly to add automatic speech recognition (ASR) functionality to user interfaces, because of the complexity of speech engine APIs, the wide variance of APIs between speech engines, and the varying computational requirements of different speech engines. For the same reason, it has also been costly to prototype new speech user interfaces (SUIs). The goal of the SpeechServer project is to lower the cost of prototyping new SUIs and adding speech to existing interfaces. The SpeechServer provides two main benefits: a simple, abstract speech API, and a separation of the computational requirements of ASR from the implementation of the SUI prototype.

Technical Discussion: The SpeechServer exposes a very simple API and set of recognition capabilities, thus breaking the link between the capabilities of the underlying recognition engine and the exposed ASR API. The SpeechServer is implemented using SPIEL, allowing the use of any speech engine that SPIEL supports. The specification of a grammar and vocabulary for the SpeechServer is also abstract, with automatic generation of data for supported ASR engines.

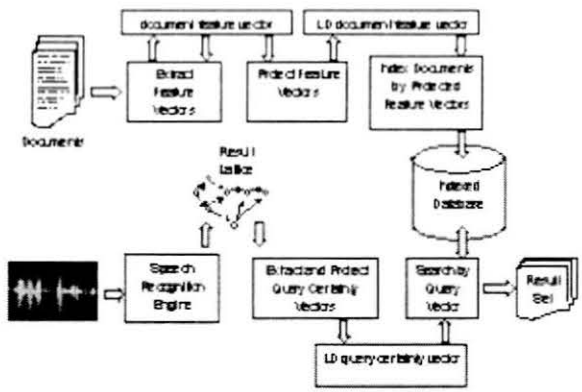
Speech recognition is a computationally expensive task. The technical requirements of a SUI vary greatly depending on the recognition task, audio environment, and target platform. Embedded/handheld applications may have tight constraints on the resources available for processing, or even require that the ASR be performed elsewhere. The SpeechServer is a distributed ASR system with a very thin client that can run on almost any system. The ASR processing is performed on a server (which may be the same system if desired), thus decoupling the SUI prototyping process from constraints imposed by prototype hardware. Therefore, SUI prototyping can proceed independently of hardware/system availability. Also, experiments with different ASR engines can be performed without change to the prototype SUI. The transport protocol between client and server is implemented on top of TCP/IP, allowing the client to be implemented in any language. MERL currently has existing clients in C++, Java, Python and ActiveX.

Collaboration: IDKen has used the SpeechServer as a component of two demonstration user interfaces in 2001-2002, including a concept prototype of "the cellphone of the future" shown at CeBIT 2001.

Future Direction: Future enhancements to the SpeechServer may include support for additional ASR engines, more complex ASR capabilities, or the addition of a complementary TTS Server.

Contact: Bret Harsham, Bent Schmidt-Nielsen, Peter Wolf **Lab:** MERL Cambridge Systems
http://www.merl.com/projects/SpeechServer/ **Project Type:** Advanced Development

SpokenQuery



SpokenQuery is technology for accessing databases using a verbal description of the desired information. It can be used to retrieve information such as web documents, music, government forms and industrial documentation using only speech. It is particularly useful in applications where hand and eyes free operation is desired. These include: call centers, information kiosks, automotive entertainment systems, Telematics, home entertainment systems, cellphone information systems, and hand held industrial systems.

Background and Objectives: For many users, search engines such as Inktomi, Google and AllTheWeb have become the primary method of locating information on the Internet. In consequence, Information Retrieval (IR) has become an important and very lucrative technology. However, the current set of search engines all require typed input, and there are many situations where a keyboard is not acceptable. Clearly, typing queries would not be acceptable while driving an automobile. The ability to access information in the automobile, on the cellphone and on PDAs combined with an interface that is hands and eyes free will enable large new markets.

The objective SpokenQuery is to enable Information Retrieval using only spoken queries. Instead of typing the query, the SpokenQuery user verbally describes the desired information. The result is a list of items that are judged to be "pertinent" to the query. Similar to current IR systems, the list is not exact but should contain a significant number of useful items.

Technical Discussion: One naïve way to implement a SpokenQuery system would be to take the text output of an open-recognition dictation system (e.g. ViaVoice, NaturallySpeaking) and feed it as input to a current IR system (e.g. Inktomi, Google). However, this implementation would not perform very well in applications that are noisy, have a far field microphone, or are speaker independent. Unfortunately, applications such as the automobile, cellphone or PDA usually have all of these problems.

Therefore MERL has developed a technique that uses more information than the simple text output of a recognizerSpokenQuery considers all the words that might have been spoken and combines this information with knowledge of the database contents (i.e. what makes sense). Our algorithm takes advantage of the low correlation between acoustics and semantics and produces good retrieval results even in extremely challenging circumstances.

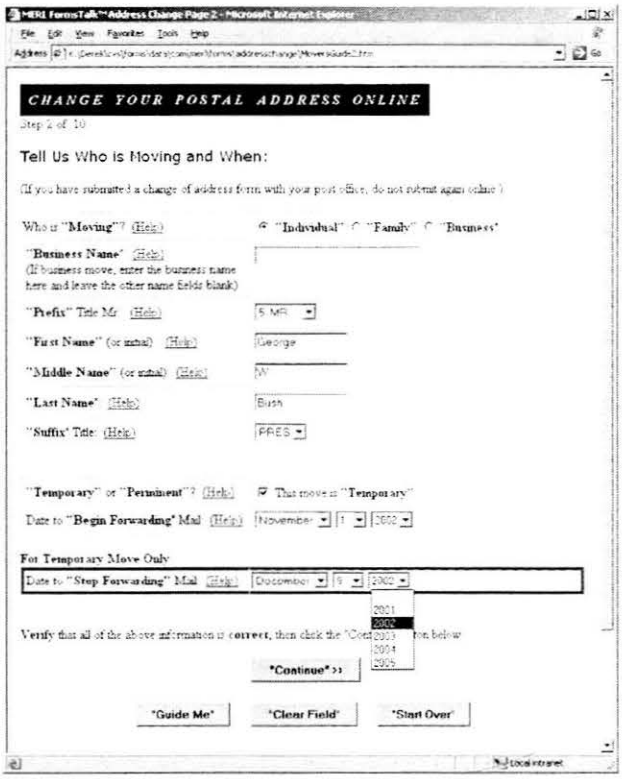
Collaboration: MediaFinder, FormsTalk.

Future Direction: Improve performance; Reduce memory and processor footprint; Produce prototype products and services for MELCO Business Units

Contact: Peter Wolf, Bhiksha Raj
http://www.merl.com/projects/SpokenQuery/

Lab: MERL Cambridge Systems
Project Type: Initial Investigation

FormsTalk



FormsTalk adds multimodal support to web form filling. Based on MERL's research in human-computer discourse and speech, this projects goals are to make innovative speech recognition and generation techniques available from within a web browser. The initial target for the system will be enhanced e-government service kiosks and web sights.

Background and Objectives: Multimodal interfaces allow different modes to augment one another, and help improve accessibility if one mode is missing. To date these systems tend to be special applications written for one domain at a time. Web based form filling promises to make web content accessible in more places than ever before (PCs, phones, cell phones, kiosks, PDAs). A side goal is to make it simpler to author web content supported on a wider range of platforms.

Technical Discussion: FormsTalk uses a java plug-in to access speech components on the local computer. It offers an active discourse agent which is not available in Microsoft's Speech .NET. This makes FormsTalk content easier to author and more natural for the user to interact with. FormsTalk supports a wide range of plugable speech recognition and generation engines including those from IBM and SpeechWorks.

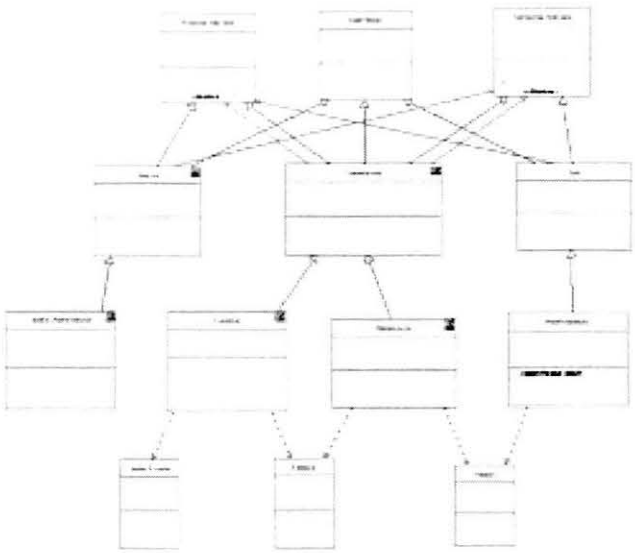
Collaboration: Working with Sentansoken's SK "Broadband Web Services" FormsTalk will provide an extendable platform for e-government systems as well as a platform for future research.

Future Direction: Both the FormsTalk platform and the e-government web pages will be evaluated, along with feedback from Sentansoken it will become the basis for future development, topics may include: dynamic content, speech processing improvements, adding support for pen ink and camera interfaces.

Contact: Bent Schmidt-Nielsen
http://www.merl.com/projects/FormsTalk/

Lab: MERL Cambridge Systems
Project Type: Research

SPIEL Toolkit



The SPIEL Toolkit is middleware for Speech applications. It enables the rapid construction of new applications and provides a common API that separates applications from the specific APIs of the speech engines. This also permits fair performance comparisons of speech recognition engines. Specified in UML, Spiel is platform and programming language independent, and especially suitable for embedded, server, and distributed applications.

Background and Objectives: Developing code that controls the speech recognition process is costly, as it must concurrently control the capture of audio data, signal processing, pattern matching, and natural language processing in a limited memory and processor footprint. Further, the probabilistic behavior of speech recognition engines requires special logging support for reproducible testing and tuning of applications.

Much of this logic is common between applications. Our objective is to identify and abstract this shared behavior as a common architecture. This greatly simplifies the production of robust and efficient applications. Since applications are not tied to a particular engine, this also enables a "best of breed" approach in choosing speech engines.

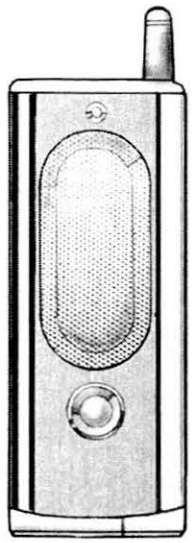
Technical Discussion: For a number of reasons, current industry standard speech middleware, such as SAPI from Microsoft and JSAPI from Sun, are not suitable for embedded, distributed or server applications. Unlike SAPI and JSAPI, SPIEL does not tie the application to an OS or language platform. Since SPIEL is specified in UML, it can be produced as C, C++ or Java, and on Windows, Linux, WINCE, and many embedded platforms.

Collaboration: In order to make SPIEL more useful for prototyping and evaluating speech applications and engines, we plan to extend it by implementing interfaces for additional speech recognition, compression and text to speech engines.

Contact: Peter Wolf
http://www.merl.com/projects/SPIEL/

Lab: MERL Cambridge Systems
Project Type: Advanced Development

ComBadge: A Voice-Operated Communications Device



The ComBadge is a two-way voice pager with a simple spoken user interface. The project encompasses the hardware, software, and user interface designs. A primary design goal has been to reduce the users' cognitive load, thus creating a communications device that is very simple and natural to use. We aim to appeal to those segments of the market where cell phone penetration is lowest, including children, the elderly, and the less-wealthy in the world.

Background and Objectives: Device costs are kept low by eliminating the display and keypad. Infrastructure costs are reduced by allowing more devices to share the available bandwidth because the messages are relatively short and the communication is asynchronous,. The spoken command set is small, so that it can be easily learned and remembered, and recognized with few errors. Familiar names are used to contact other users by having each user add customized voice name tags for other ComBadges.

Technical Discussion: Speech recognition, audio compression, and radio transmission do not overlap, thereby reducing the peak power demand and extending battery life. Compression need not occur in real-time, which permits the use of a slower processor and/or a better compression algorithm. Inexpensive bandwidth intended for data, rather than voice, can be used at all stages of the network. Message delivery could be accomplished over the Internet.

Asynchronous messaging also has advantages for users. The device can be very small, since it does not need to reach from mouth to ear. Users are less aware of dead spots in network coverage and are less irritated by network outages due to overloading, since these conditions produce delays rather than dropped calls. Furthermore, the ComBadge is less intrusive because users determine when they want to listen and respond to messages.

Collaboration: We are working very closely with the speech applications group at MERL Cambridge Systems.

Future Direction: We are exploring using the ComBadge for group messaging — a spoken distribution list allows the same message to be delivered to several users. Delivering voice messages to machines as well as people is an open-ended opportunity. Initially, we are looking at using ComBadge to control household devices and as a voice portal to a wide variety of Internet services.

Contact: Joe Marks
http://www.merl.com/projects/ComBadge/

Lab: MERL Cambridge Research
Project Type: Research

Color Figures

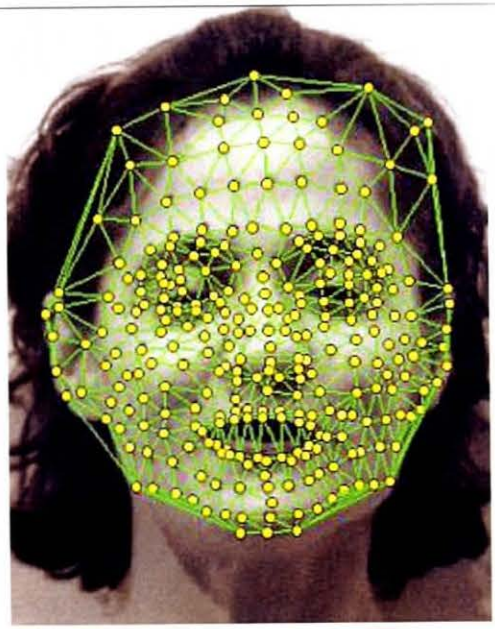


Figure 1 – see page 52

Figure 2 – see page 60

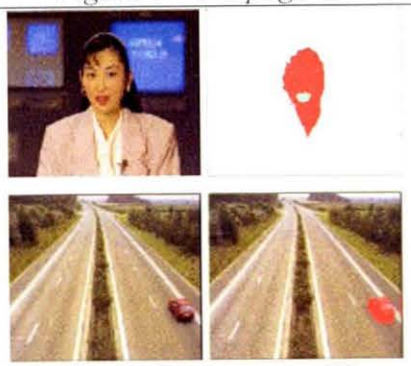


Figure 3 – see page 63

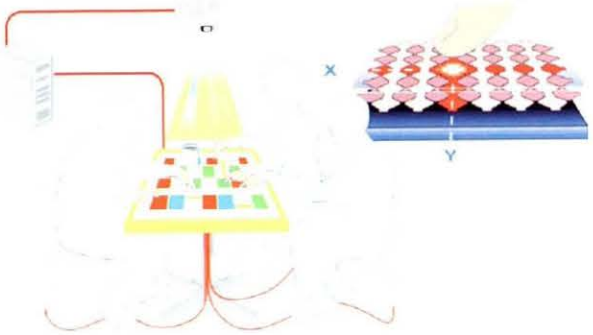


Figure 4 – see page 68

Figure 5 – see page 71

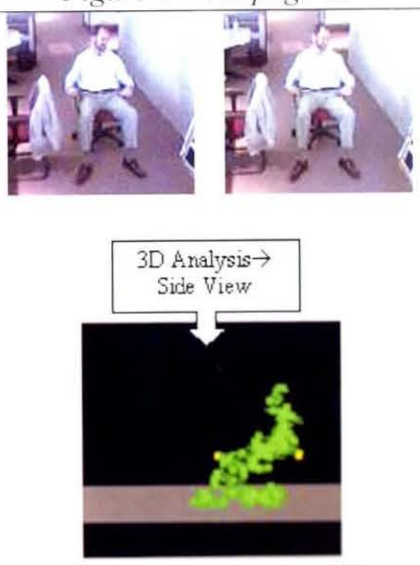


Figure 6 – see page 80

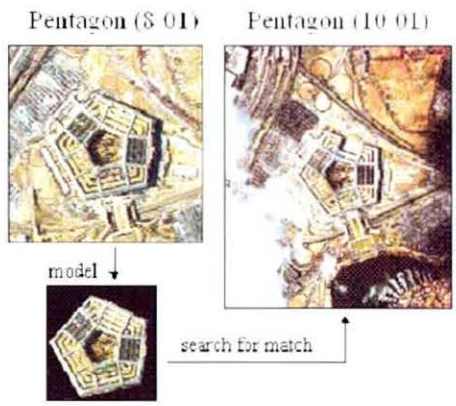


Figure 7 – see page 85

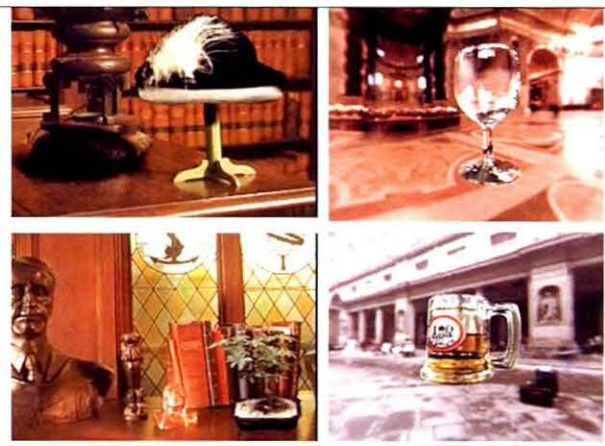


Figure 8 – see page 106

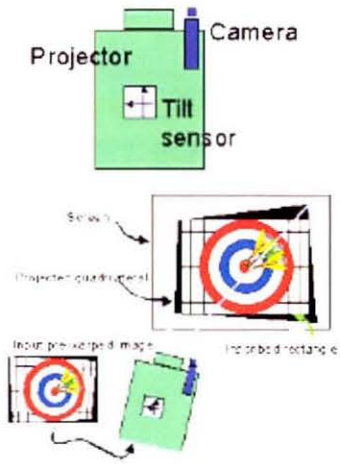


Figure 9 – see page 107

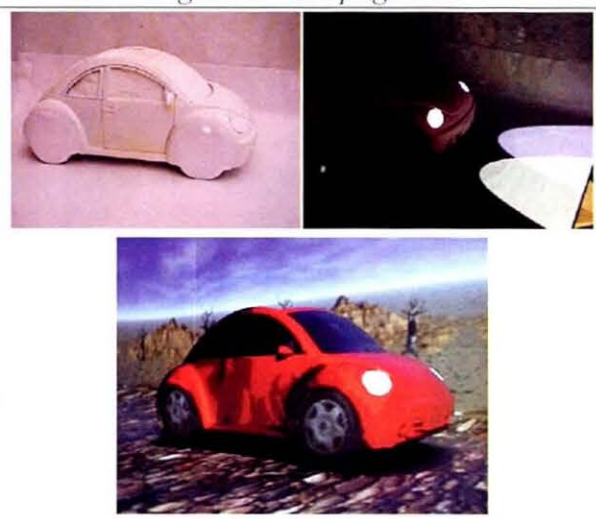


Figure 10 – see page 108

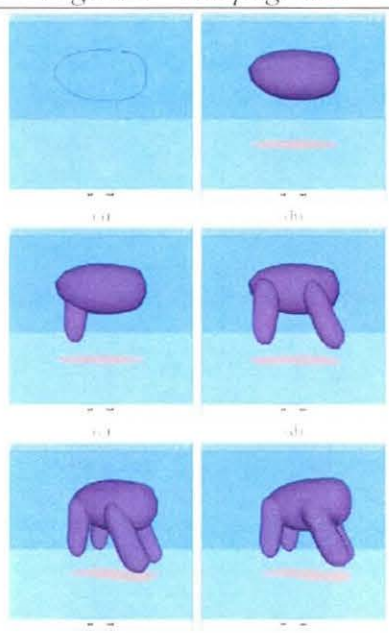


Figure 11 – see page 109

Figure 12 – see page 119